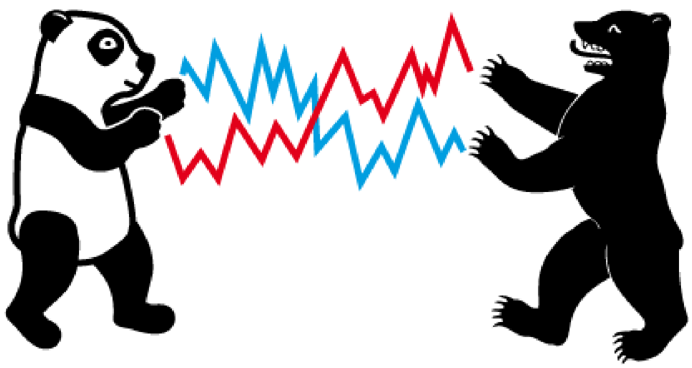
IRTG 1792 Discussion Paper 2019-018

Phenotypic convergence of cryptocurrencies

Daniel Traian Pele [*]
Niels Wesselhöfft [*2]
Wolfgang K. Härdle [*2] [*3] [*4]
Michalis Kolossiatis [*5]
Yannis Yatracos [*2]



[*] The Bucharest University of Economic Studies, Romania
[*2] Humboldt-Universität zu Berlin, Germany
[*3] Singapore Management University, Singapore
[*4] Charles University, Czech Republic
[*5] University of Cyprus, Cyprus

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Phenotypic convergence of cryptocurrencies

Daniel Traian Pele[a], Niels Wesselhöfft[b], Wolfgang K. Härdle[c], Michalis Kolossiatis[d], Yannis Yatracos[b]

[a]*Department of Statistics and Econometrics, Faculty of Cybernetics, Statistics and Economic Informatics, The Bucharest University of Economic Studies, 010371, Bucharest, Romania, E-mail: danpele@ase.ro*

[b]*Humboldt-Universität zu Berlin, IRTG 1792, Dorotheenstr. 1, 10117 Berlin, Germany, E-mail: wesselhn@hu-berlin.de*

[c]*Humboldt-Universität zu Berlin, IRTG 1792, Dorotheenstr. 1, 10117 Berlin, Germany School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899 Faculty of Mathematics and Physics, Charles University, Ke Karlovu 3, 121 16 Prague, Czech Republic*

[d]*Department of Mathematics and Statistics, University of Cyprus, Nicosia, 1678, Cyprus Central Bank of Cyprus, Nicosia 1076, Cyprus, E-mail: kolossiatis.michalis@ucy.ac.cy*[1]

Abstract

The aim of this paper is to prove the phenotypic convergence of cryptocurrencies, in the sense that individual cryptocurrencies respond to similar selection pressures by developing similar characteristics. In order to retrieve the cryptocurrencies phenotype, we treat cryptocurrencies as financial instruments (*genus proximum*) and find their specific difference (*differentia specifica*) by using the daily time series of log-returns. In this sense, a daily time series of asset returns (either cryptocurrencies or classical assets) can be characterized by a multidimensional vector with statistical components like volatility, skewness, kurtosis, tail probability, quantiles, conditional tail expectation or fractal dimension. By using dimension reduction techniques (Factor Analysis) and classification models (Binary Logistic Regression, Discriminant Analysis, Support Vector Machines, K-means clustering, Variance Components Split methods) for a representative sample of cryptocurrencies, stocks, exchange rates and commodities, we are able to classify cryptocurrencies as a new asset class with unique features in the tails of the log-returns distribution. The main result of our paper is the complete separation of the cryptocurrencies from the other type of assets, by using the Maximum Variance Components Split method. More, we observe a divergent evolution of the cryptocurrencies species, compared to the classical assets, mainly due to the tails behaviour of the log-returns distribution. The codes used here are available via www.quantlet.de.

Keywords: cryptocurrency, genus proximum, differentia specifica, classification, multivariate analysis, factor models, phenotypic convergence, divergent evolution
JEL Classification: C14, C22, C46, C53, G32

[1]The information and views set out in this article are those of the authors and do not necessarily reflect the official opinion of the Central Bank of Cyprus.

1. Introduction

Cryptocurrencies, served as a new digital asset, have attracted much attention from both investors and academics. Along with this growing popularity, the market capitalization of cryptocurrencies is increasing substantially. Thus, according to a recent report (Transparency Market Research, 2018), the total capitalization for cryptocurrencies market was around US$574.3 mn in the year 2017 and is expected to reach US$6702.1 mn by the end of 2025.

Most articles focus on Bitcoin (BTC), as it is considered the first decentralized cryptocurrency, which has the largest capitalization from its beginning till now. An extensive review of the literature regarding the Bitcoin can be found in Corbet et al. (2019). Table 1 lists a synthesis of the empirical findings regarding the statistical properties of the cryptocurrencies, compared to classical assets.

Given preceding results from the literature, our contribution to the studies dealing with the cryptocurrencies market is mostly empirical, proving the complete separation of cryptocurrencies from the other assets and the validity of phenotypic convergence in case of cryptocurrencies.

Our approach is quite different from the existing literature, as most of the reviewed paper are using a low-dimensional approach when trying to differentiate cryptocurrencies from classical assets, by using either market risk indicators or long memory indicators. By using a multidimensional approach and taking into account various statistics describing the tail, moment and memory behaviour of the time series of daily log-returns, we use a genus-differentia approach[2] in order to find the specific difference (*differentia specifica*) of the cryptocurrencies, regarded as financial instruments (*genus proximum*).

[2] A genus-differentia approach describes a species first by a broader category, the genus, then that species is distinguished from other items in that category by a specific difference, i.e. *per genus et differentiam* (Blackburn, 2008).

Through means of dimensionality reduction techniques (Factor Analysis), we prove that most of the variation among cryptocurrencies, stocks, exchanges rates and commodities can be explained by three factors: the tail factor, the moment factor and the memory factor.
Our results add to the findings from literature by showing that the most important factor which differentiates the cryptocurrencies from classical assets is the tail behaviour of the log-returns distribution, as proven by classification techniques: Binary Logistic Regression, Discriminant Analysis, Support Vector Machines and Variance Components Split. In other words, by looking at cryptocurrencies as financial instruments (proximal genus), their specific difference is the tail behaviour of the log-returns distribution. This finding is confirmed by the classical Factor Analysis, performed on a static basis and also by using the expanding window approach, where the assets universe is seen in an evolutionary dynamic.

The main result of our paper is the complete separation of the cryptocurrencies from the other type of assets, by using the Maximum Variance Components Split method; the other methods (Binary Logistic Regression, Discriminant Analysis, Support Vector Machines,K-means clustering) provide an almost complete separation.

Another important result is the discovery of a phenotypic convergence of cryptocurrencies, compared to the classical assets. In biology, the phenotype of an organism (Mahner and Kary, 1997) is the set of the organism's observable characteristics (i.e. morphology, developmental processes, biochemical and physiological properties). Phenotypic convergence or convergent evolution can be defined as "the appearance of similar phenotypes in distinct evolutionary lineages" (Washburn et al., 2016).

If the assets universe is regarded as an ecosystem, then we can construct an analogy with the biological concepts. Thus, the phenotype of an asset can be determined by some statistical features of the time series of prices or returns. In this paper, the phenotypic convergence refers to the fact that individual cryptocurrencies tend to develop over time certain similar characteristics that make them fully distinguishable from the classical assets, i.e. the cryptocurrencies tend to behave like a homogenous group, with certain characteristics that individualize them in the assets ecosystem. By using an expanding window approach, we are able to show that the cryptocurrencies have a convergent dynamic in relationship to the classical assets and this convergence is driven mainly by the tail behaviour of the log-returns distribution. More, the cryptocurrencies as a species exhibit a divergent evolution in relation to classical assets. Originated from biology, the concept of divergent evolution refers to the accumulation of differences between related populations, leading to speciation (Rieseberg et al., 2004). Divergent evolution is typically exhibited when two populations are exposed to different selective factors, driving their adaptation to the environment (Bergstrom and Dugatkin, 2016). A related analysis can be found in ElBahrawy et al. (2017), where the cryptocurrencies market is seen as an evolutive system, with several characteristics which are preserved over time. According to ElBahrawy et al. (2017), the evolution of the cryptocurrencies market has been ruled by "neutral" forces, i.e. no cryptocurrency has shown any strong selective advantage over the other.

The paper is subsequently organized as follows: the second section describes the methodology used, including Factor Analysis, Binary Logistic Regression, Support Vector Machines (SVM), K-means clustering, Variance Components Split (VCS) methods and the Evolutive divergence; the third section describes the datasets and interprets the results of the classification; the fourth section describes the phenotypic convergence, while the last section concludes. The codes used to obtain the results in this paper are available via `www.quantlet.de`.

Table 1: Empirical findings on the cryptocurrencies market

Authors	Assets	Sample	Findings
Dyhrberg (2016a)	BTC, USD/EUR, USD/GBP, FTSE index.	2010-2015, daily data.	BTC can act as a hedge between UK equities and the USD.
Dyhrberg (2016b)	BTC, Federal funds rate, USD/EUR, USD/GBP, FTSE index, Gold futures, Gold cash.	2010-2015, daily data.	BTC is somewhere in between a currency (USD) and a commodity (Gold).
Bariviera et al. (2017)	BTC, USD/EUR, USD/GBP.	2011-2017, daily data. 2013-2016, intraday data.	BTC presents large volatility and long-range memory (Hurst exponent higher than 0.5). BTC standard deviation is 10 times greater than other currencies.
Baur et al. (2018)	BTC, Federal funds rate, USD/EUR, USD/GBP, FTSE index, Gold futures, Gold cash.	2010-2015, daily data.	BTC returns are not a hybrid of Gold and USD returns.
Caporale et al. (2018)	BTC, LTC, Ripple, Dash	2013-2017, daily data.	The four cryptocurrencies exhibit persistence (Hurst exponent higher than 0.5), yet the persistency degree changes over time.
Härdle et al. (2018)	BTC, XRP, LTC, ETH, gold and S&P500	2016-2018, daily data.	BTC, XRP, LTC, ETH exhibit higher volatility, skewness and kurtosis compared to Gold and S&P500 daily returns.
Henriques and Sadorsky (2018)	BTC and five exchange traded funds (ETFs): US equities (SPY), US bonds (TLT), US real estate (VNQ), Europe and Far East equities (EFA), and Gold (GLD).	2011-2017, daily data.	BTC can be a substitute for Gold in an investment portfolio, achieving a higher risk adjusted return.
Jiang et al. (2018)	BTC	2010-2017, daily data.	Long-term memory and high degree of inefficiency ratio exists in the BTC market.
Klein et al. (2018)	BTC, CRIX index, Gold, Silver, crude oil prices for the West Texas Intermediate (WTI), the S&P500 index, MSCI World and the MSCI Emerging Markets 50 index.	2011-2017, daily data.	BTC returns have the highest mean and standard deviation.
Selmi et al. (2018)	BTC, Gold, Brent crude oil	2011-2017, daily data.	Both BTC and Gold would serve the roles of a hedge, a safe haven and a diversifier for oil price movements.
Stosic et al. (2018)	Top 119 cryptocurrencies.	2016-2017, daily data.	Collective behaviour of the cryptocurrency market.
Takaishi (2018)	BTC, GBP/USD	2014-2016, intraday data	The 1-min return distribution of BTC is fat-tailed, with high kurtosis; BTC time series exhibits multifractality.
Urquhart (2016)	BTC	2010-2016, daily data.	Hurst statistic indicates strong anti-persistence (values lower than 0.5).
Wei (2018)	456 different cryptocurrencies.	2017, daily data.	Lower volatility for liquid cryptocurrencies. Illiquid cryptocurrencies exhibit strong return anti-persistence in the form of a low Hurst exponent.
Zhang et al. (2018)	70 % of cryptocurrencies market.	2013-2018, daily data.	Cryptocurrencies exhibit heavy tails, quickly decaying returns autocorrelations, slowly decaying autocorrelations for absolute returns, strong volatility clustering, leverage effects, long-range dependence, power-law correlation between price and volume.
Borri (2019)	BTC, ETH, LTC, XRP, Gold Bullion, the CBOE volatility index (VIX), the S&P400 commodity chemicals index, and the S&P500 index.	2017-2018, daily data.	Cryptocurrencies exhibit large and volatile return swings, and are riskier than most of the other assets.

2. Methodology

The methodology used in this paper has four layers: first, we study the statistical properties of the daily log-returns of the selected assets and we estimate the components of a multidimensional vector describing the behaviour of the time series of assets' daily log-returns. Second, we apply data dimension reduction and orthogonalization methods (Factor Analysis) in order to retain the orthogonal factors which maximize the explained variance. Third, we employ classification techniques (Binary Logistic Regression, Discriminant analysis, Support Vector Machine, K-means clustering, Variance Components Split methods) to obtain the most influential factors discriminating between the cryptocurrencies and the classical assets. Fourth, we prove the validity of the phenotypic convergence, showing that cryptocurrencies pose specific characteristics, allowing them to differentiate over time from the classical assets.

2.1. Taxonomy variables

According to Aristotle, the definition of a species consists of genus proximum and differentia specifica (Parry and Hacker, 1991). In order to properly define cryptocurrencies in terms of their genus proximum and differentia specifica, we need an initial dataset of variables that have the statistical power to differentiate between the cryptocurrencies and the classical assets (stocks, exchange rates and commodities).

Before introducing the multidimensional dataset used for taxonomy, we set the following notation:

- n: the number of assets in the dataset;
- t: the time index, $t \in \{1, \ldots, T\}$, where T is the time of the last record in the dataset;

- $P_{i,t}$: the closing price for asset i in day t, with $i = 1 \ldots n$, $t = 1 \ldots T$;
- $R_{i,t} = \log P_{i,t} - \log P_{i,t-1}$: the daily log-return for asset i in day t, with $i = 1 \ldots n$, $t = 1 \ldots T$;
- $R = (R_{i,t})_{\substack{i=1\ldots n \\ t=1\ldots T}} \in M(n,T)$: the initial matrix of the assets' daily log-returns;
- p: the number of variables used for taxonomy.

The multidimensional dataset used for taxonomy is the matrix $X_t = (x_{it,k})_{\substack{i=1\ldots n \\ k=1\ldots p}} \in M(n,p)$, whose components are detailed below, estimated for the time interval [1, t], with $t \in \{t_0, .., T\}$, where $t_0 = [T/3]$ (the integer part of $T/3$).
First, we took into account the central moments of the log-returns distribution, through the following parameters:

- variance: $\sigma^2_{it} = E\left\{(R_i - \mu_{i,t})^2\right\}$;
- skewness: $Skewness_{it} = E\left\{(R_i - \mu_{i,t})^3\right\}/\sigma^3_{it}$;
- kurtosis: $Kurtosis_{it} = E\left\{(R_i - \mu_{i,t})^4\right\}/\sigma^4_{it}$.

Second, we estimated the following parameters of the α-stable distribution, fitted to daily log-returns, in order to capture some scaling properties:

- the tail parameter: $Stable_{\alpha_{it}} \in$ (0,2], lower values indicating heavier tails;
- the scale parameter: $Stable_{\gamma_{it}} > 0$.

The α-stable distributions are a well-known class of distributions used in financial modelling (Rachev and Mittnik, 2000), capturing the fat tails and the asymmetries of the real-world log-returns distributions. The α-stable parameters were estimated using the empirical characteristic function method, following Koutrouvelis (1980, 1981). For computational efficiency, we used the fast parallel α-stable distribution function evaluation and parameter estimation, through the

Matlab library *libstable* (Julián-Moreno et al., 2017).

Third, we estimated the quantiles and the conditional tail expectations for the distribution of log-returns, in order to capture the tail behaviour:

- quantiles: $Q_{\alpha;it}$, with $\alpha \in \{0.005, 0.01, 0.025, 0.05, 0.95, 0.975, 0.99, 0.995\}$;
- conditional left tail expectation: $CTE_{\alpha,it}(R_{it}) = E[R_{it}|R_{it} < Q_{\alpha;it}]$, for $\alpha \in \{0.005, 0.01, 0.025, 0.05\}$;
- conditional right tail expectation: $CTE_{\alpha,it}(R_{it}) = E[R_{it}|R_{it} > Q_{\alpha;it}]$, for $\alpha \in \{0.95, 0.975, 0.99, 0.995\}$.

From a market risk perspective, the left tail quantiles can be assimilated to Value-at-Risk, the conditional left tail expectation can be regarded as Expected Shortfall, while the conditional right tail expectation can be seen as the Expected Upside. Fourth, we estimated the following parameters, in order to capture the memory properties:

- first order autocorrelation of the time series of daily log-returns: $\rho_{it}(1)$;
- Hurst exponent: H_{it}. The Hurst exponent (Hurst, 1951) of the time series of daily log-returns is estimated based on the discrete second-order derivative in the wavelet domain (Istas and Lang, 1997).

Our multidimensional dataset can be seen as a tensor $\mathscr{X} \in \mathbb{R}^{n \times p \times T'}$, where n is the number of assets, $p = 23$ is the number of variables and $T' = T - t_0$ is the number of time points.

2.2. Factor Analysis

The most popular methods used to synthesize and extract relevant information from large datasets are Principal Components Analysis (PCA) and Factor Analysis (FA)(Bartholomew, 2011). In this paper, we use Factor Analysis to extract the main factors explaining the variation in the initial dataset, the reason for this being that PCA is a linear combination of variables, while FA is a measurement model of a latent variable. The aim of Factor Analysis is to explain the outcome of the p variables in the data matrix X using fewer variables, the so-called factors (Härdle and Simar, 2012). The orthogonal factor model is given by:

$$X = QF + U + \mu, \tag{1}$$

with the following notations:

- X is the initial matrix of p variables;
- F are the common k factors ($k << p$);
- Q is a matrix of the non-random loadings of the common factors F;
- U is a matrix of the random specific factors;
- μ is the vector of the means of initial p variables.

It also holds that the random vectors F and U are unobservable and uncorrelated.

In our paper, the initial factor pattern is extracted using the principal component method, followed by a VARIMAX rotation to insure orthogonality of the factors. The Factor Analysis is applied on the entire dataset X_T, the p initial variables being estimated for the entire time period $[1, T]$. The p-dimensional dataset X_T is then projected on the k-dimensional space defined by the k orthogonal factors, in order to observe a separation of the components of the assets.

2.3. Assets Classification

In order to find *genus proximum* and *differentia specifica* of the cryptocurrencies, we are using several classification techniques, detailed below.

2.3.1. Binary Logistic Regression

The Binary Logistic Regression model quantifies the performance of each of the orthogonal factors extracted through the Factor Analysis to discriminate between the cryptocurrencies and classical assets. Thus, we are estimating the following family of models:

$$P(Y_i = 1) = \frac{\exp(\beta_{0j} + \beta_{1j} F_{ji})}{1 + \exp(\beta_{0k} + \beta_{1j} F_{ji})}, \tag{2}$$

where $Y_i = 1$ for cryptocurrencies, $Y_i = 0$ for classical assets, and $F_j, j \in \{1, \ldots, k\}$ are the k orthogonal factors retrieved through the factor analysis. Based on the explanatory power and the significance of model 2, we can derive the most important factors contributing to the *differentia specifica* of cryptocurrencies. As a performance measure for model 2, we are using $\tilde{R}^2$ (Nagelkerke, 1991), where:

$$\tilde{R}^2 = \frac{1 - \left\{\frac{L(\mathbf{0})}{L(\widehat{\boldsymbol{\beta}})}\right\}^{\frac{2}{n}}}{1 - \{L(\mathbf{0})\}^{\frac{2}{n}}}. \tag{3}$$

In (3), $L(0)$ is the likelihood of the intercept-only model, $L(\widehat{\boldsymbol{\beta}})$ is the likelihood of the full model, and $\widehat{\boldsymbol{\beta}}$ is the vector of Maximum Likelihood estimated parameters.

2.3.2. Discriminant Analysis

The aim of discriminant analysis is to classify one or more observations into *a priori* known groups, minimising the error of misclassification (Härdle and Simar, 2012). Formally, Linear Discriminant Analysis (LDA) assumes that the input dataset is multivariate Normal: $X_i \sim N(\mu_i, \Sigma)$, where X_i belong to class ω_i. The goal is to project samples X onto a line $Z = w^\top X$, where we select the projection that maximizes the standardized separability of the means over all directions. Specifically, we maximize the normalized, squared distance in the means of the classes

$$w^* = \arg\max_{w} \frac{|w^\top(\mu_i - \mu_j)|^2}{s_i^2 + s_j^2}, \tag{4}$$

$$s_i^2 = \sum_{x_i \in \omega_i} (w^\top x_i - w^\top \mu_i)^2 = w^\top S_i w, \tag{5}$$

giving the Linear Discriminant of Fisher (Fisher, 1936):

$$w^* = S_W^{-1}(\mu_i - \mu_j), \;\; S_W = S_i + S_j. \tag{6}$$

Quadratic Discriminant Analysis (QDA) follows the same procedure, but for $X_i \sim N(\mu_i, \Sigma_i)$ belong to the class ω_i, one can relax the condition of equality of covariance matrices by $\Sigma_i \neq \Sigma_j$, $i \neq j$, allowing for a non-linear classifier.

2.3.3. Support Vector Machines

Support Vector Machines (SVM) is a data classification technique, its goal being to produce a model which predicts target values based on a set of attributes (Cristianini and Shawe-Taylor, 2000). The goal is to find a projection that maximizes margin in a hyperplane of the original data, without any parametric assumptions on the underlying stochastic process. The support vectors are determined via a quadratic optimization problem i.e. given a training data set D with n samples and 2 dimensions $D = (X_1, Y_1), \ldots (X_n, Y_n)$, $X_i \in \mathbb{R}^2$, $\quad Y_i \in [0, 1]$, the aim is to find a hyperplane that maximizes the margin:

$$\min_{w,b} \frac{1}{2}\|w\|^2, \;\text{ s.t. } \quad Y_i \left(w^\top X_i + b\right) \geq 1, i = 1, \ldots, n. \tag{7}$$

2.3.4. K-means Clustering Algorithm

This clustering method was first popularized by (MacQueen, 1967) (in that paper, the author acknowledges a couple of other researchers that independently used that method around the same time). The aim is to allocate each observation of a data set in one of $k \in \mathbb{N}$ clusters, where k is predefined, so as to minimize the within-cluster sums of squares. In brief, the algorithm proceeds as follows:

i. Take k data points and set them as the cluster centres.

ii. Iteratively, for each data point, assign it to the cluster which centre is closer to the data point (the Euclidean distance is usually used, but other distance metrics have been proposed). Update the cluster centre for the selected cluster.

iii. Repeat until convergence (*i.e.* the allocations do not change).

2.3.5. Variance Components Split methods: MVCS, GMVCS

These methods have goals to separate, respectively, the components of a structure like the types of assets herein or the types of Iris flowers, and clusters defined as the components of a mixture distribution. They are based on an unusual variance decomposition in between-group variations (Yatracos, 1998, 2013). To describe the sample version of the decomposition, let $X_1, \ldots, X_n$ be *i.i.d.* random variables; $X_{(j)}$ is the j-th order statistic, $1 \leq j \leq n$. Consider the groups $X_{(1)}, \ldots, X_{(i)}$ and $X_{(i+1)}, \ldots, X_{(n)}$ with averages, respectively, $\bar{X}_{[1,i]}$ and $\bar{X}_{[i+1,n]}$, $i = 1, ..., n-1$, then

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n-1}\frac{i(n-i)}{n^2}(\bar{X}_{[i+1,n]} - \bar{X}_{[1,i]})(X_{(i+1)} - X_{(i)}). \tag{8}$$

The summands in the right side of (8) measure between-groups variations. The standardized sample variance components

$$W_i = W_i(X_1, \ldots, X_n) = \frac{i(n-i)}{n}\frac{(\bar{X}_{[i+1,n]} - \bar{X}_{[1,i]})(X_{(i+1)} - X_{(i)})}{\sum_{i=1}^{n}(X_i - \bar{X})^2},\ i = 1, \ldots, n-1, \tag{9}$$

indicate the relative contribution of the groups $X_{(1)}, ..., X_{(i)}$ and $X_{(i+1)}, ..., X_{(n)}$ in the sample variability. Index

$$\mathcal{I}_n = \max\{W_i, i = 1, \ldots, n-1\} \tag{10}$$

determines two potential clusters or parts of a structure and is based on averages and inter-point distances. When $\mathcal{I}_n = W_j$, these clusters are $\tilde{\mathcal{C}}_1 = \{X_{(1)}, \ldots, X_{(j)}\}, \tilde{\mathcal{C}}_2 =$

$\{X_{(j+1)}, \ldots, X_{(n)}\}$. The observed $\mathcal{I}_n$-value is significant at α-level for the normal model when it exceeds the critical value $[-\ln(-\ln(1-\alpha)) + \ln n]/n$ (Yatracos, 2009); $\alpha = 0.05$ is used herein.

When $\mathcal{X}$ is the n by r data matrix of r-dimensional observations, $\mathbf{X}_j$ is the j-th row of $\mathcal{X}, j = 1, \ldots, n$. The coefficients of the orthogonal projection of $\mathcal{X}$ along the unit norm r-row vector $\mathbf{a}$ are $\mathcal{X}\mathbf{a} = (\mathbf{X}_1\mathbf{a}, \ldots, \mathbf{X}_n\mathbf{a})$. The split in the sorted values of $\mathcal{X}\mathbf{a}$ where

$$\mathcal{I}_{\mathcal{X}}(\mathbf{a}) = \max\{W_i(\mathbf{X}_1\mathbf{a}, \ldots, \mathbf{X}_n\mathbf{a}); i = 1, ..., n-1\} \tag{11}$$

is attained, determines *along* $\mathbf{a}$ the groups $\tilde{\mathcal{C}}_{\mathcal{X},1}(\mathbf{a})$ and $\tilde{\mathcal{C}}_{\mathcal{X},2}(\mathbf{a})$ in the $\mathcal{X}$-rows which are potential clusters and parts of a structure. For example, if for the data herein $\tilde{\mathcal{C}}_{\mathcal{X},1}(\mathbf{a})$ consists of rows 1-14, the cryptocurrencies (a component) among the assets (the structure) are completely separated along $\mathbf{a}$.

The Maximum Variance Component Split (MVCS) method compares known components of a structure, *e.g.* cryptocurrencies herein, with data splits for a set of unit projection directions $\mathcal{D}_M$ usually determined by M positive equidistant angles of $[0, \pi]$; *e.g.* when $r = 2$ and $M = 3$ the angles used are $\pi/3, 2\pi/3, \pi$. When one of the data split along projection direction $\mathbf{a}$ coincides with a component of the structure we have complete separation of this component along $\mathbf{a}$.

A set of projection directions $\mathcal{D}_M$ can be

$$(\Pi_{l=1}^{r} cos\theta_l,\ sin\theta_1 \Pi_{l=2}^{r} cos\theta_l, ...,\ sin\theta_{r-1} cos\theta_r,\ sin\theta_r), \tag{12}$$

where θ_l takes values in $\{\frac{m\pi}{M}, m = 1, ..., M\}, l = 1, ..., r$.

The method is computationally intensive for large r and M values, thus it may be used on subsets of the $\mathcal{X}$-columns. The importance of a subset S of $\mathcal{X}$-columns in the separation of a structure's component is measured by the number N_S of projection directions (12) completely separating the component. Indications for the importance of a specific column c in S in the separation of the same component are obtained by comparing N_S with the number of projection directions N_{S-c} separating the component when c is left out and also by comparing all $N_{S-c}, c \in S$. Similar indications of importance can be used for subgroups of S-columns.

The Global Maximum Variance Component Split (GMVCS) along all projection vectors $\mathcal{D}$, to be obtained from $\max\{\mathcal{I}_{\mathcal{X}}(\mathbf{a}), \mathbf{a} \in \mathcal{D}\}$ that is called "the index" determines 2 clusters. In practice, its approximation is obtained using $\mathcal{D}_M$. The splitting of these clusters may continue or not (Yatracos, 2013).

2.3.6. Expanding window modelling

For observing the assets dynamic, we are using an expanding window approach, allowing to distinguish the evolution of the clusters. In fact, for $t \in \{t_0, .., T\}$, where $t_0 = [T/3]$, the p-dimensional dataset X_t is projected on the k-dimensional space defined by the main factors extracted through the Factor Analysis applied on the dataset X_T. By using this projection instead of a time-varying factor model, we are avoiding situations like changes in factors loadings, causing inconsistencies over time.

3. Data and Results

Our dataset is a combination of cryptocurrencies and classical assets (commodities, exchange rates and stocks), covering the time period 10/20/2014 - 10/16/2018 (1006 trading days), for $n = 544$ assets (see Table 2). The reason for choosing this time span for the analysis is that before 2015 the liquidity in the cryptocurrencies market had been relatively low, their total market capitalization being less than US$16 billion (Feng et al., 2018).

Table 2: Dataset

Type of Asset	Number of Assets	Source
Cryptocurrencies	14	Coinmarketcap
Stocks	497	Bloomberg
Exchange rates	13	Bloomberg
Commodities	20	Bloomberg

The first component of the dataset contains a representative sample of cryptocurrencies; the cryptocurrencies selected to be part of this analysis are the components of the CRIX Index (Härdle et al., 2018), sourced from https://coinmarketcap.com/. The CRyptocurrency IndeX is a benchmark for the cryptocurrencies market, being real-time computed by the Ladislaus von Bortkiewicz Chair of Statistics at Humboldt University Berlin, Germany. The 15 components of the CRIX index (see Appendix A) account for 90% of the total cryptocurrencies market capitalization, as seen in Figure 1. In our analysis, the USDT cryptocurrency is discarded due to the fact that it is an atypical digital currency, having very little variation around the reference value of US$1.

The second component of the dataset contains a sample of the most traded commodities indexes, according to Bloomberg (see Appendix A).

The third component of the dataset contains a sample of the most liquid exchange rates, according to Bloomberg (see Appendix A).

The fourth component of the dataset contains the constituents of the S&P500 Index, recorded at October 19th 2018. The number of constituents of S&P500 Index is 505, however in our dataset only 497 of them are listed, i.e. those stock with valid data for the entire time period analysed (the complete list of the stocks included in the analysis can be found here).

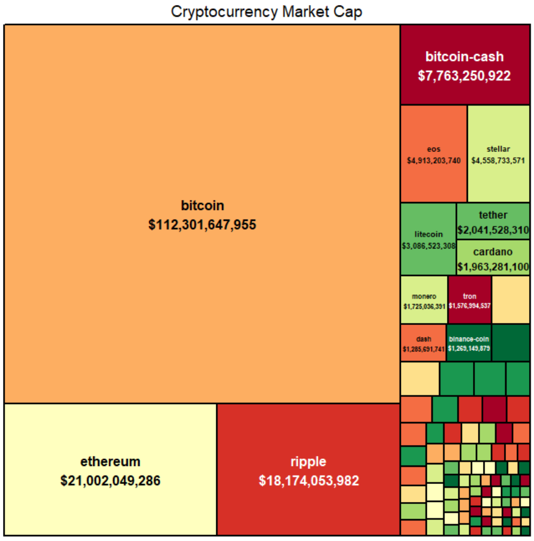


Figure 1: Cryptocurrencies market capitalization (US$) at 10/19/2018.

Mkt_cryptos.

3.1. Factor Analysis

Factor analysis is a classical method used to find latent variables or factors among observed variables, by grouping variables with similar characteristics together. Performing the factor analysis requires, in general, three steps:

i. Estimation of the correlation matrix for all the variables, shown in Figure 2.

ii. Extraction of the factors from the correlation matrix, based on the correlation coefficients of the variables.

iii. Factor rotation, in order to maximize the relationship between the variables and some of the factors.

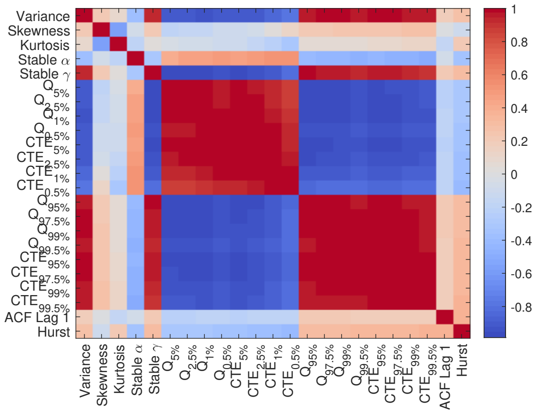


Figure 2: Correlation matrix. SFA_cryptos

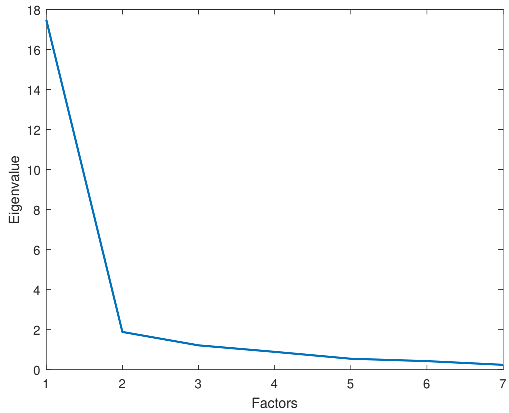


Figure 3: Scree plot. SFA_cryptos

Based on the eigenvalues criteria, we can select those factors for which the eigenvalue is higher than 1 (see Figure 3, where the scree plot is shown). According to this criteria, three factors were selected, accounting for 89.6% of the total variance.

In order to test the sampling adequacy of the factor analysis, we are using the Kaiser-Meyer-Olkin (KMO) test; this test should be greater than 0.5 for a satisfactory factor analysis (Tabachnick and Fidell, 2013).

The overall KMO test is computed using the following formula:

$$KMO = \frac{\sum_i \sum_{i \neq j} r_{ij}^2}{\sum_i \sum_{i \neq j} r_{ij}^2 + \sum_i \sum_{i \neq j} u_{ij}^2} \tag{13}$$

where $R = (r_{ij})_{\substack{i=1\ldots n \\ j=1\ldots n}}$ is the correlation matrix and $U = (u_{ij})_{\substack{i=1\ldots n \\ j=1\ldots n}}$ is the partial covariance matrix (Cerny and Kaiser, 1977, Kaiser, 1974).

The individual KMO test is computed using the formula:

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} u_{ij}^2} \tag{14}$$

In fact, the KMO measure represents the proportion of the variance in the input variables that might be caused by underlying factors (Kaiser, 1981). The overall

KMO value is 0.903, pointing out that the factor analysis is suitable for structure detection (see Table 3). For the factor rotation, we used the VARIMAX method, which outputs orthogonal factors, also minimizing the number of variables that have high loadings on each factor.

Based on the rotated factors pattern, the following conclusions can be drawn (see also Figure 4):

i. The first factor – **the tail factor**, accounting for 76.1% of the total variance, is highly correlated with the following parameters: the tail parameter alpha and the scale parameter gamma of the stable distribution, the lower and upper quantiles of the distribution of log-returns, the conditional tail expectations and the variance of log-returns.

ii. The second factor – **the moment factor**, accounting for 8.2% of the total variance, is highly correlated with the skewness and kurtosis of the distribution of log-returns.

iii. The third factor – **the memory factor**, accounting for 5.3% of the total variance, is highly correlated with the Hurst exponent and the first order autocorrelation coefficient of log-returns.

Table 3: Kaiser's Measure of Sampling Adequacy

Variable	KMO measure	Variable	KMO measure
$Variance$	0.970	$Q_{99.5\%}$	0.893
$Skewness$	0.540	$CTE_{0.5\%}$	0.877
$Kurtosis$	0.510	$CTE_{1\%}$	0.892
$Stable_{\alpha}$	0.935	$CTE_{2.5\%}$	0.928
$Stable_{\gamma}$	0.861	$CTE_{5\%}$	0.884
$Q_{0.5\%}$	0.923	$CTE_{95\%}$	0.878
$Q_{1\%}$	0.932	$CTE_{97.5\%}$	0.898
$Q_{2.5\%}$	0.921	$CTE_{99\%}$	0.884
$Q_{5\%}$	0.915	$CTE_{99.5\%}$	0.874
$Q_{95.5\%}$	0.899	$\rho(1)$	0.713
$Q_{97.5\%}$	0.948	H	0.862
$Q_{99\%}$	0.925	Overall KMO	0.903

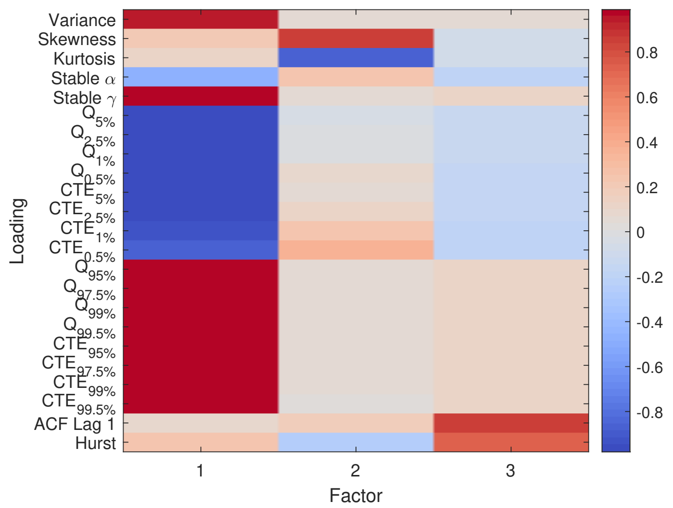


Figure 4: Loadings of the three factors. SFA_cryptos

Based on the data revealed in Table 4, one can synthesize few characteristics of the cryptocurrencies that differentiate them from the other assets:

- The cryptocurrencies have higher variance of the log-return's distribution, compared to the classical assets.
- The cryptocurrencies exhibit the presence of heavy tails, as indicated by high values of quantiles and conditional tail expectations, i.e. the cryptocurrencies have higher propensity for risk.
- The first order autocorrelation is positive in the case of cryptocurrencies, while all the other assets have negative first order autocorrelation, for the analysed time period.
- Bitcoin is closer to classical stocks and commodities, in terms of the tail factor, i.e. its risk profile can be considered at the border between the classical assets and cryptocurrencies.

Table 4: Assets profile based on the average values of the initial variables

Factor	**Variable**	**Cryptos**	**Stocks**	**Commodities**	**Exchange Rates**	**Bitcoin**
Tail	$\sigma^2 \cdot 10^3$	7.880	0.280	0.370	0.030	1.500
factor	$Stable_\alpha$	1.436	1.703	1.753	1.759	1.319
	$Stable_\gamma \cdot 10^3$	36.760	8.730	9.850	3.170	16.020
	$Q_{0.5\%}$	−0.255	−0.056	−0.052	−0.015	−0.139
	$Q_{1\%}$	−0.215	−0.044	−0.043	−0.013	−0.113
	$Q_{2.5\%}$	−0.148	−0.032	−0.034	−0.010	−0.086
	$Q_{5\%}$	−0.113	−0.024	−0.026	−0.008	−0.063
	$Q_{95\%}$	0.133	0.024	0.027	0.008	0.059
	$Q_{97.5\%}$	0.198	0.030	0.035	0.010	0.082
	$Q_{99\%}$	0.285	0.040	0.045	0.013	0.109
	$Q_{99.5\%}$	0.383	0.050	0.056	0.015	0.139
	$CTE_{0.5\%}$	−0.326	−0.077	−0.072	−0.022	−0.184
	$CTE_{1\%}$	−0.278	−0.063	−0.060	−0.018	−0.152
	$CTE_{2.5\%}$	−0.217	−0.048	−0.046	−0.014	−0.123
	$CTE_{5\%}$	−0.172	−0.038	−0.038	−0.011	−0.098
	$CTE_{95\%}$	0.233	0.035	0.040	0.011	0.092
	$CTE_{97.5\%}$	0.307	0.043	0.049	0.013	0.116
	$CTE_{99\%}$	0.411	0.055	0.065	0.016	0.147
	$CTE_{99.5\%}$	0.499	0.067	0.080	0.019	0.175
Moment	$Skewness$	0.973	−0.508	0.285	−1.223	−0.283
factor	$Kurtosis$	20.351	12.922	20.716	33.992	8.583
Memory	$\rho(1) \cdot 10^3$	40.630	−2.160	−13.180	−11.450	16.640
factor	H	0.567	0.509	0.533	0.514	0.565

Based on the factors estimated through the factor analysis, one can map the cryptocurrencies and the classical assets, in order to derive some clustering effect.

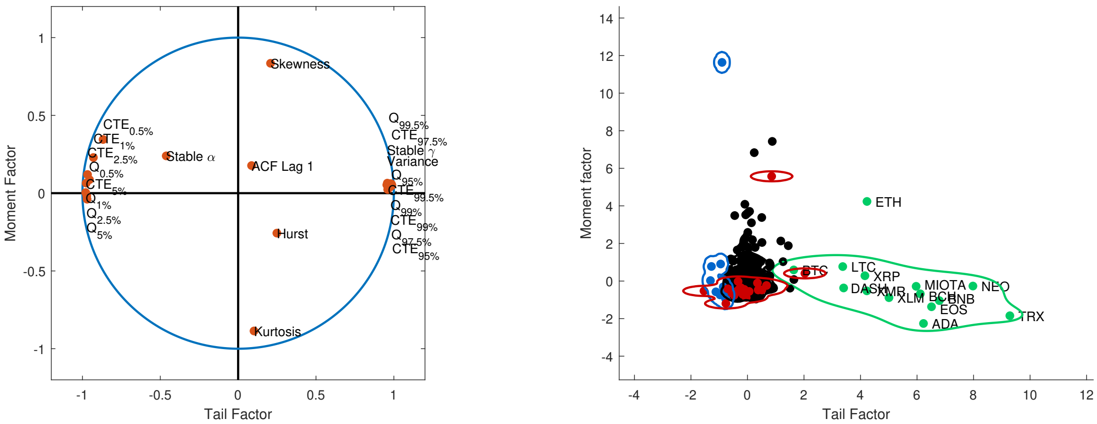


Figure 5: Loadings (left) and scores (right) based on tail and moment factor.
SFA_cryptos

Figures 5 to 7 map the cryptocurrencies and the classical assets; the colour code is the following: green: cryptocurrencies, black: stocks, red: commodities, blue: exchange rates. Also, a 95% confidence region is estimated, based on the Bivariate Kernel Density (in green).

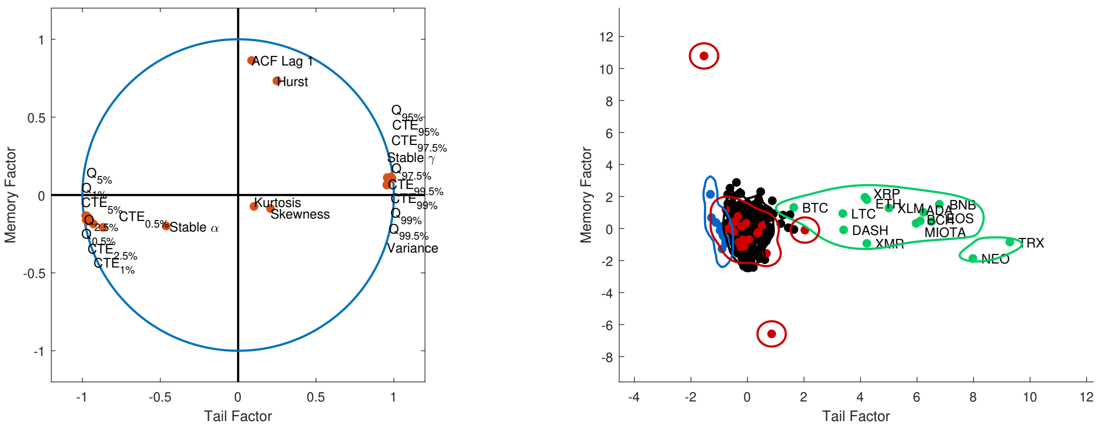


Figure 6: Loadings (left) and scores (right) based on tail and memory factor.
SFA_cryptos

As shown in Figure 5 and Figure 6, there is a clear separation between cryptocurrencies and classical assets, mainly due to the first factor, the tail factor, while the memory and moment factor are of subliminal importance (see Figure 7).

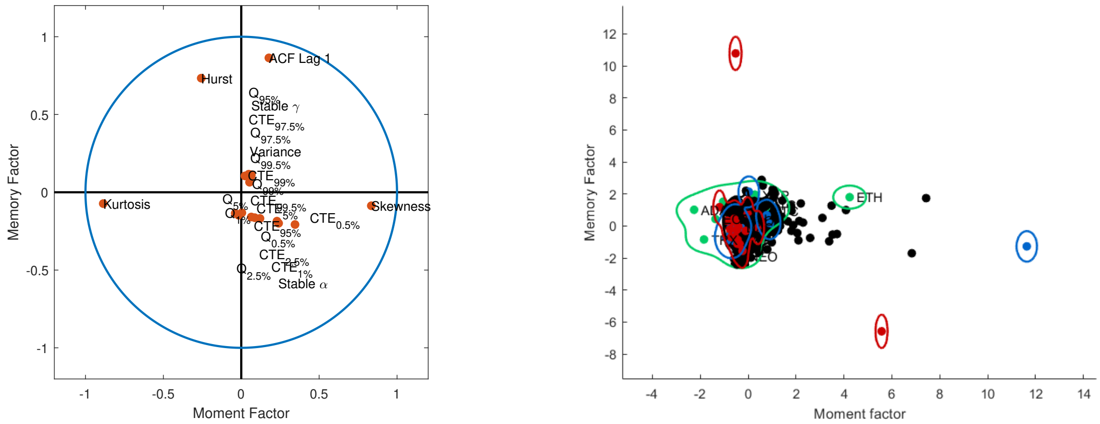


Figure 7: Loadings (left) and scores (right) based on moment and memory factor.

SFA_cryptos

3.2. Assets classification

In this section, we list the results of the models presented in Section 2.3, in order to assess the ability of the factors produced through the factor analysis to discriminate between cryptocurrencies and classical assets.

First, for each of the three factors we estimated the Binary Logistic model:

$$P(Y_i = 1) = \frac{\exp(\beta_{0j} + \beta_{1j} F_{ji})}{1 + \exp(\beta_{0k} + \beta_{1j} F_{ji})}, \tag{15}$$

where $Y_i = 1$ for cryptocurrencies, $Y_i = 0$ for classical assets, and $F_j, j \in \{1, 2, 3\}$ are the orthogonal factors retrieved through the factor analysis.

Table 5 lists the estimated β_{1j} of the Binary Logistic Regression model (15), with the performance measure defined by equation (3).

As seen in Table 5, the most important factor regarding the separation between the cryptocurrencies and the classical assets is the tail factor, while the other two factors have no significant influence.

Table 5: Estimates of model (15)

Exogenous factor	Factor 1	Factor 2	Factor 3
Esimated β_1	4.398**	-3.729	-3.692
	(2.086)	(-0.606)	(0.314)
$\tilde{R}^2$	0.958	0.015	0.024

Note: Standard errors in parentheses; ** denotes significance at 95% confidence level.

Second, we employed Discriminant Analysis and Support Vector Machines on the space defined by the two first factors (tail and memory). Figure 8 presents the classification results using Discriminant Analysis. Both linear and quadratic classifiers have a very good classification power, the only cryptocurrency which is misclassified being the Bitcoin (see Table 4 for Bitcoin's profile).

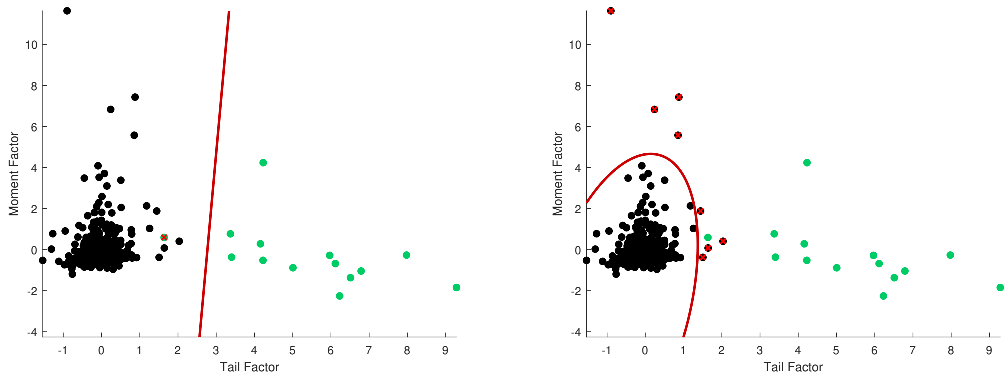


Figure 8: Discriminant Analysis: linear(left) and quadratic (right). Green dots denote the cryptocurrencies, while the black dots denote the other assets; the dots highlighted in red are cases of misclassification. SFA_cryptos

The same conclusion can be drawn by looking at the results of the Support Vector Machines non-linear classifier, according to which all the cryptocurrencies are correctly classified using the tail factor and the moment factor (see Figure 9).

The k-means clustering algorithm (MacQueen, 1967) was also used for k =

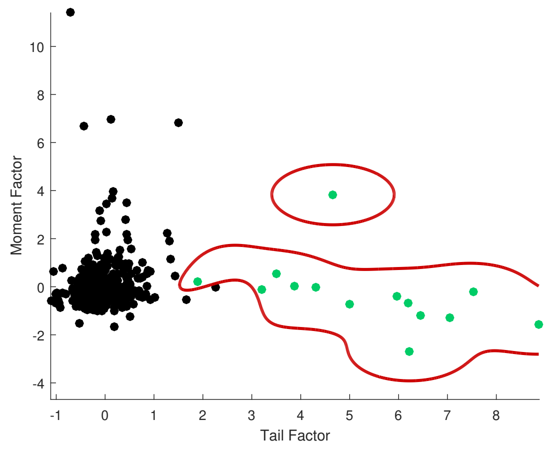


Figure 9: Support Vector Machines. Green dots denote the cryptocurrencies, while the black dots denote the other assets. 95% confidence region is in red. SFA_cryptos

$2, 3, 4$, the results showing that this method does not provide perfect classification[3] for any of the assets.

We used $k = 2, 3, 4$ and for all three cases, no perfect classification was obtained for any of the assets. When $k = 2$, one cluster consists of four observations (two stocks, one exchange rate and one commodity), while all the other assets are in the second cluster. Also, for $k = 3$, the first cluster consists of two observations (one exchange rate and one commodity), the second cluster consists of one cryptocurrency, ten stocks and one commodity, while all the other assets are in the third cluster. For $k = 4$, the first cluster consists of three stocks, the second cluster consists of two observations (one exchange rate and one commodity), the third cluster consists of thirty eight assets (two cryptocurrencies, thirty three stocks, two exchange rates and one commodity) and all the remaining assets are in the last cluster.

The results when applying the MVCS method, where the goal is to achieve sep-

[3]Perfect classification is the case when a specific component is completely separated by the rest. In other words, all the members of that component, and only those, are in one cluster.

aration of the components of the assets, are in accordance, but even stronger than those of Binary Logistic Regression, Discriminant Analysis and Support Vector Machines. In order to apply MVCS method, we are considering the following term structure: the assets-data are regarded as a matrix $X_T = (x_{iT,k})_{\substack{i=1...n \\ k=1...p}} \in M(n,p)$, with $p = 23$ columns, representing the variables used for taxonomy, and $n = 544$, representing the assets. To be concise, the 23 columns are considered to be ordered, using the following order: $Variance$, $Skewness$, $Kurtosis$, $Stable_\alpha$, $Stable_\gamma$, $Q_{0.5\%}$, $Q_{1\%}$, $Q_{2.5\%}$, $Q_{5\%}$, $Q_{95.5\%}$, $Q_{97.5\%}$, $Q_{99\%}$, $Q_{99.5\%}$, $CTE_{0.5\%}$, $CTE_{1\%}$, $CTE_{2.5\%}$, $CTE_{5\%}$, $CTE_{95\%}$, $CTE_{97.5\%}$, $CTE_{99\%}$, $CTE_{99.5\%}$, $\rho(1)$, H. The following notations are used for the MVCS method: M are the positive equidistant angles of $[0; \pi]$, S is a specific subset of the columns, N_S is the number of projection directions giving perfect classification when S is used, P_S is the corresponding percentage of these directions, while $minI, maxI$ are the minimum and the maximum index I value for perfect classification, respectively. The critical value for significance of the index for $\alpha = 5\%$ and $n = 544$ is 0.017.

In the following, we are presenting the results of the MVCS method for perfect classification of cryptocurrencies from the other assets, as it was found that for all three other structures (stocks, exchange rates and commodities), none of the combinations of M and S presented below provided perfect classification.

Table 6 shows the results of the MVCS method, which gave a perfect classification, when splitting the columns into disjoint subsets 1-12 and 13-23. Due to processing power constraints, projection directions (12) are used only for $M = 3, 6$. The number of projection directions used is M^{d-1}, with d, respectively, 12 and 11.

Table 7 shows the results of the MVCS method, when splitting the columns into disjoint subsets 1-6 and 7-12.

Regarding columns 13-18 and 19-23, the values $M = 24$ and 36 were are also used, however no projection direction gave perfect classification for cryptocurrencies.

Table 6: Results of the MVCS method

M	S	N_S	P_S	$minI$	$maxI$
3	1-12	172	0.097%	0.043	0.165
6	1-12	631968	0.17%	0.036	0.212
3	13-23	0	0	n/a	n/a
6	13-23	0	0	n/a	n/a

Table 7: Results of the MVCS method, columns 1-6 and 7-12

M	S	N_S	P_S	$minI$	$maxI$
3	1-6	2	0.82%	0.092	0.095
6	1-6	266	3.42%	0.051	0.178
9	1-6	133	0.23%	0.056	0.181
12	1-6	4263	1.71%	0.043	0.181
15	1-6	998	0.13%	0.047	0.182
18	1-6	18890	1%	0.0454	0.183
3	7-12	0	0	n/a	n/a
6	7-12	0	0	n/a	n/a
9	7-12	10	0.017%	0.060	0.097
12	7-12	170	0.068%	0.067	0.130
15	7-12	131	0.017%	0.059	0.165
18	7-12	328	0.017%	0.054	0.163

We can conclude that the most important columns for complete separation are columns 1-12, and in particular columns 1-6 (as can be seen from P_S, for the same value of M). The projected values for all the assets, using columns 1-12, on the projection direction that provided the largest index value are shown in Figure 10.

The projection direction that gave the largest index value is: (0.106, 0, 0, -0.061, 0.211, 0, 0.141, 0.162, -0.187, -0.217, -0.75, 0.5); the index value in this

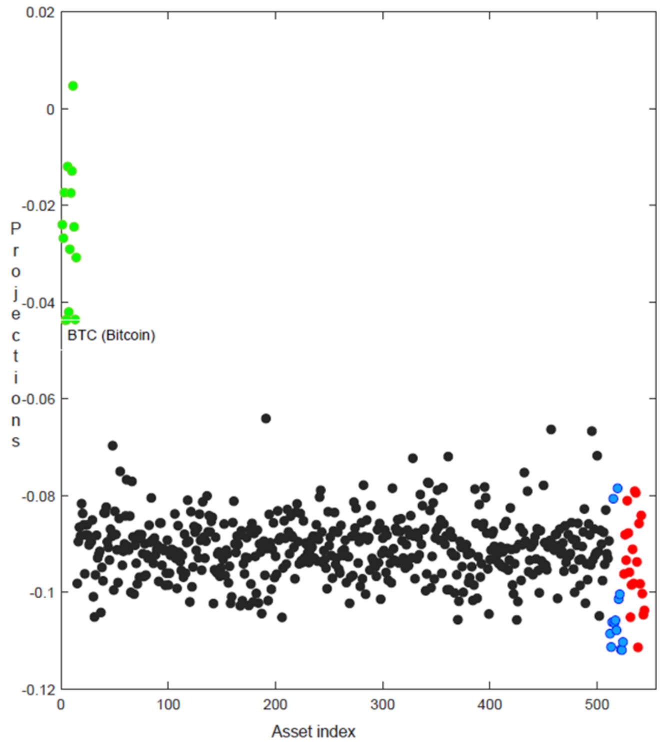


Figure 10: Projections of a subset of the data (the first 12 columns) on the projection direction that gave the largest index value

direction is 0.213. For columns 1-12 and $M = 6$, in total 631968 (out of 362797056 tried) projection directions gave perfect classification and all provided statistically significant index values for the normal model. Finally, the first six columns are used, as they are deemed the most important, according to the above. Then, the MVCS method is applied to all six quintets (these quintets are derived by omitting one of the six columns). Again, higher values of M are used, the results being reported in Table 8.

From the above, one can see that the least important columns are the second and the third (Skewness and Kurtosis), while the most important columns are the sixth ($Q_{0.05}$) and then the fourth ($Stable_{\alpha}$). As mentioned before, perfect classification is obtained only for cryptocurrencies, while the other assets have in-

Table 8: Results for cryptocurrencies, all leave-one-out quintets of columns 1-6

M	S	N_S	P_S	$minI$	$maxI$
18	1,2,3,4,5	8	0.0004%	0.045	0.149
18	1,2,3,4,6	361	0.0190%	0.051	0.179
18	1,2,3,5,6	333	0.0176%	0.080	0.107
18	1,2,4,5,6	2040	0.1001%	0.045	0.183
18	1,3,4,5,6	1856	0.0982%	0.045	0.183
18	2,3,4,5,6	990	0.0524%	0.051	0.181
24	1,2,3,4,5	21	0.0003%	0.050	0.152
24	1,2,3,4,6	1247	0.0157%	0.050	0.179
24	1,2,3,5,6	589	0.0074%	0.081	0.109
24	1,2,4,5,6	5644	0.0709%	0.040	0.184
24	1,3,4,5,6	4550	0.0570%	0.043	0.184
24	2,3,4,5,6	2328	0.0292%	0.054	0.179
32	1,2,3,4,5	1105	0.0033%	0.054	0.153
32	1,2,3,4,6	2316	0.0069%	0.048	0.180
32	1,2,3,5,6	1049	0.0031%	0.070	0.115
32	1,2,4,5,6	16072	0.0480%	0.035	0.184
32	1,3,4,5,6	11018	0.0328%	0.042	0.184
32	2,3,4,5,6	5326	0.0159%	0.046	0.181

distinct behaviour.[4] This result is in line with the conclusions obtained trough the other classification techniques used above (Binary Logistic Regression, Discriminant Analysis and Support Vector Machines), MVCS method showing that cryptocurrencies behaves like a totally different species in the assets universe.

From an Aristotelian point of view (i.e. *genus proximum et differentia speci-*

[4]Similar results are obtained when applying MVCS in the first 16 columns with $M = 3$.

fica), we can conclude that cryptocurrencies are financial instruments (proximal genus) whose specific difference is the tail behaviour of the distribution of daily log-returns. In other words, based on the tail factor profile, we can conclude that a random asset is likely to be a cryptocurrency if it has the following properties: very long tails of the log-returns distribution (in terms of the left and right quantile and the conditional tail expectation), high variance, high value of the α-stable scale parameter and value of the α-stable tail index close to 1.

4. Phenotypic convergence

For observing the assets dynamic, we are using an expanding window approach, allowing to distinguish the evolution of the clusters. In fact, for $t = t_0, \ldots, T$, the p-dimensional dataset is projected on the k-dimensional space defined by the main factors extracted through the Factor Analysis applied on the dataset X_T. By using this projection instead of a time-varying factor model, we are avoiding situations like changes in factors loadings, causing inconsistencies over time.

In order to derive the dynamics of the assets' universe, we used an expanding window approach, described below:

- First, the 23-dimensional dataset is estimated for the time interval $[1, t_0] = [10/10/2014, 02/19/2016]$.
- Second, the time window is extended on a daily basis, up to T=10/16/2018 and for each step in time, the dataset is projected on the 2-dimensional space defined by the tail factor and the moment factor, estimated for the entire time period.

Figure 11 presents a snapshot of the evolution of the assets universe using the

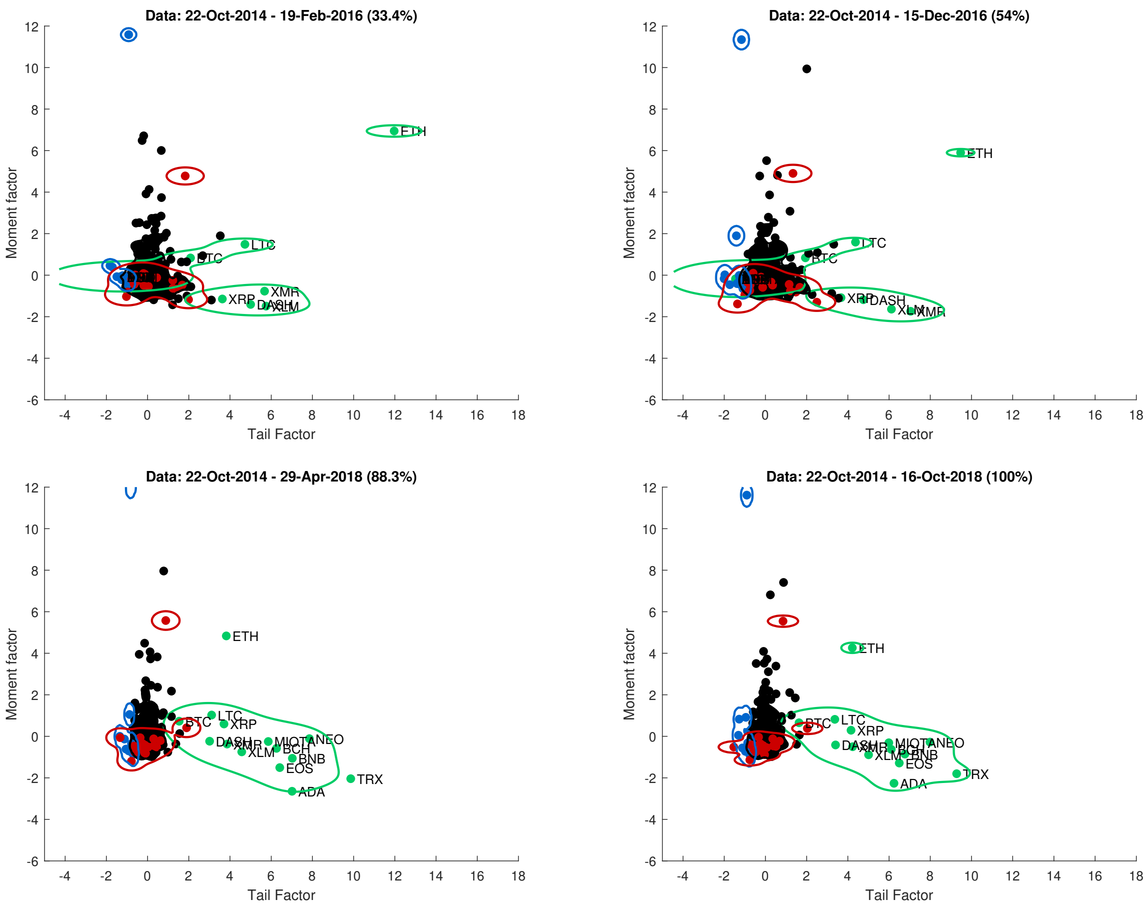


Figure 11: The evolution of the assets universe using the expanding window approach. The colour code is the following: green: cryptocurrencies, black: stocks, red: commodities, blue: exchange rates. DFA_cryptos

expanding window approach[5]. Looking at the evolution of the assets universe, it appears that the behaviour of cryptocurrencies can be described by the concepts of phenotypic convergence and divergent evolution. These concepts refer to the fact that individual cryptocurrencies tend to develop over time similar characteristics (phenotypic convergence) that make them fully distinguishable from the classical

[5]The daily evolution of the assets universe, for the period 02/19/2016-10/16/2018, is depicted in the video *Crypto_movie*, attached to this paper as supplementary material.

assets (divergent evolution).

In order to test this behaviour, we are using the Likelihood Ratio associated to model (2), estimated using the expanding window approach previously described. The Likelihood Ratio for this model can be defined as:

$$LR(\widehat{\boldsymbol{\beta}}) = -2(\log L(\widehat{\boldsymbol{\beta}}) - \log L(\widehat{\boldsymbol{\beta}_s})), \tag{16}$$

where $L(\widehat{\boldsymbol{\beta}_s})$ is the likelihood of a saturated model that fits perfectly the sample, while $L(\widehat{\boldsymbol{\beta}})$ is the likelihood of the estimated model. In the languange of Binary Logistic Regression, the Likelihood Ratio from the equation (16) is called deviance (Hosmer and Lemeshow, 2010) and is a measure of model goodness-of-fit, with large values indicating models with poor classification power. The deviance is always positive, being zero only for the perfect fit.

In order to derive the statistical significance of the classification, we compare the Likelihood Ratios of the estimated model and of the intercept-only model. Thus, we compute the difference of the likelihood ratios:

$$D = [LR(\widehat{\boldsymbol{\beta}}) - LR(0)], \tag{17}$$

where asymptotically $D \sim \chi^2(1)$, $LR(0)$ being the likelihood ratio of the intercept-only model. In fact, we are estimating m models, where $m = T - t_0 - 1 = 971$ and for each model we report the Likelihood Ratio (Figure 12) and the p-value associated to equation (17) (see Figure 13).

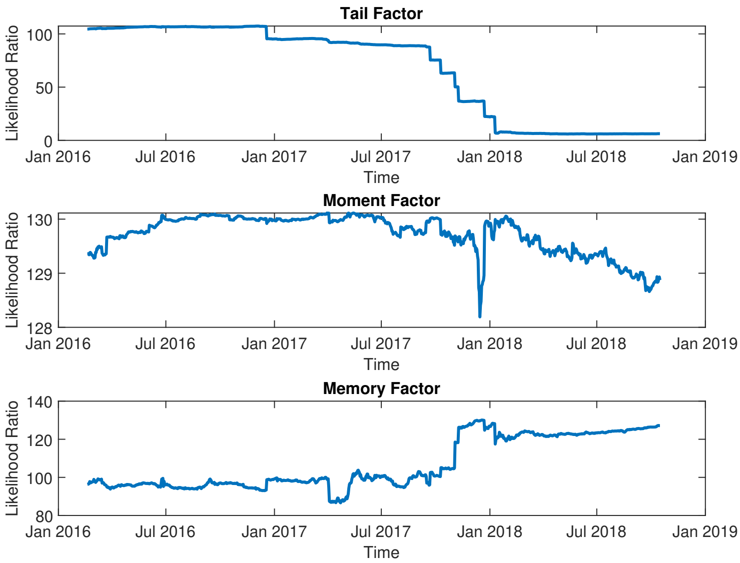


Figure 12: Likelihood Ratios for model (2), estimated on the time period 02/19/2016-10/16/2018, using an expanding window approach. CONV_cryptos

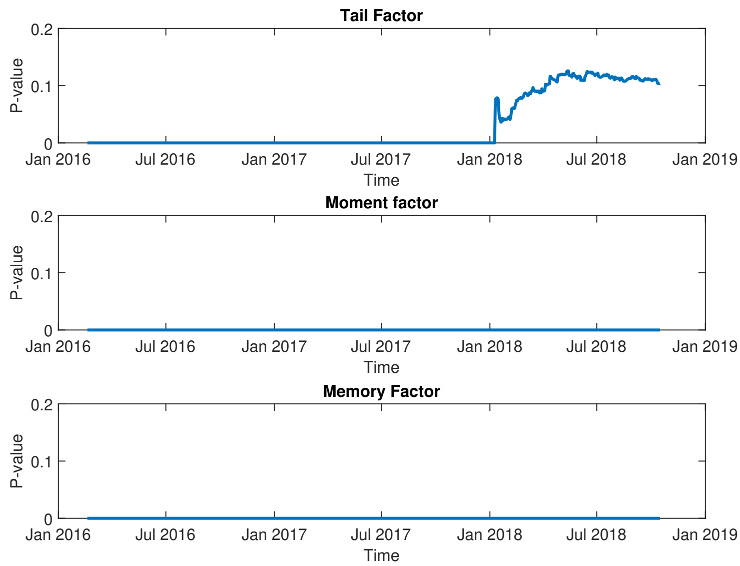


Figure 13: p-values for equation (17), estimated on the time period 02/19/2016-10/16/2018, using an expanding window approach. CONV_cryptos

By examining the evolution of the p-values, we can observe that the shift in significance for the tail-factor-based model is recorded on January 2018, when the cryptocurrencies market collapsed, after the historical maximum of Bitcoin from December 2017.

The most important implication of this finding is the validity of phenotypic convergence among cryptocurrencies: in their evolution, the individual cryptocurrencies have developed similar characteristics (heavier tails, higher volatility, higher propensity to extreme negative returns), that differentiate them from the classical assets and position them as a new, different species in the ecosystem of financial instruments.

5. Conclusions

In this paper we applied a genus-differentia approach in order to discriminate between the cryptocurrencies and the classical assets, like stocks, exchange rates and commodities. By using a multidimensional approach and taking into account various indicators describing the market risk behaviour, the tail behaviour and the long-memory characteristics of the time series of daily log-returns, we found the specific difference of cryptocurrencies, regarded as financial instruments (proximal genus).

Through the means of dimensionality reduction techniques and classification techniques, we proved that most of the variation among the cryptocurrencies, stocks, exchanges rates and commodities can be explained by three factors: the tail factor, the moment factor and the memory factor. Our analysis revealed that the main difference between cryptocurrencies and classical assets, in terms of properties of the distribution of daily log-returns, is the tail behaviour, both in the left and in the right tail of the distribution. The moments of the distribution and the time-series memory are of subliminal importance for discriminating between cryptocurrencies and classical assets.

Based on the tail factor profile, we can conclude that a random asset is likely to be a cryptocurrency if it has the following properties: very long tails of the log-returns distribution (in terms of the left and right quantile and the conditional tail expectation), high variance, high value of the α-stable scale parameter and value of the α-stable tail index closer to 1.

Moreover, the cryptocurrencies are completely separated by the other types of assets, as proved by Maximum Variance Components Split method.

From the point of view of the risk analysts and regulators, the non-linear classification techniques applied on the factors extracted provide proficient results in order to discriminate between the cryptocurrencies and the other assets.

The added value of our research is the study of the cryptocurrencies universe using the concepts of phenotypic convergence and divergent evolution. Through the means of an expanding window approach, we are able to depict the evolutionary dynamics of cryptocurrencies universe and show how the clusters formed by projecting the multidimensional dataset on the main factors converge over time.

By looking at the assets universe as a complex ecosystem, we are able to conclude that the cryptocurrencies exhibit both a phenotypic convergence (individual cryptocurrencies develop similar characteristics over time) and a divergent evolution, as different species, compared to the classical assets.

Acknowledgements

Financial support from the Deutsche Forschungsgemeinschaft, Germany via IRTG 1792 "High Dimensional Non Stationary Time Series", Humboldt-Universität zu Berlin, Germany, is gratefully acknowledged.

Appendix A - List of commodities, exchange rates and cryptocurrencies used in the analysis

Table A.1: List of commodities

Nr.crt.	Commodity	Symbol
1	WTI Crude oil	USCRWTIC Index
2	Natural Gas	NGUSHHUB Index
3	Brent oil	EUCRBRDT Index
4	Unleaded Gasoline	RBOB87PM Index
5	ULS Diesel	DIEINULP Index
6	Live cattle	SPGSLC Index
7	Lean hogs	HOGSNATL Index
8	Wheat	WEATTKHR Index
9	Corn	CRNUSPOT Index
10	Soybeans	SOYBCH1Y Index
11	Aluminum	LMAHDY Comdty
12	Copper	LMCADY Comdty
13	Zinc	ZSDY Comdty
14	Nickel	CKEL Comdty
15	Tin	JMC1DLTS Index
16	Gold	XAU Curncy
17	Silver	XAG Curncy
18	Platinum	XPT Curncy
19	Cotton	COTNMAVG Index
20	Cocoa	MLCXCCSP Index

Table A.2: List of exchange rates

Nr. crt.	Symbol	Denomination	Name
1	EUR	EUR/USD	Euro
2	JPY	JPY/USD	Japanese Yen
3	GBP	GBP/USD	Great Britain Pound
4	CAD	CAD/USD	Canada Dollar
5	AUD	AUD/USD	Australia Dollar
6	NZD	NZD/USD	New Zealand Dollar
7	CHF	CHF/USD	Swiss Franc
8	DKK	DKK/USD	Danish Krone
9	NOK	NOK/USD	Norwegian Krone
10	SEK	SEK/USD	Swedish Krone
11	CNY	CNY/USD	Chinese Yuan Renminbi
12	HKD	HKD/USD	Hong Kong Dollar
13	INR	INR/USD	Indian Rupee

Table A.3: CRIX components at 10/19/2018

Coin	Symbol	Name	Market Cap (in $K)
1	BTC	Bitcoin	114,953,322
2	ETH	Ethereum	21,665,771
3	XRP	Ripple	19,035,356
4	BCH	Bitcoin Cash	7,975,384
5	EOS	EOS	5,005,087
6	XLM	Stellar	4,633,717
7	LTC	Litecoin	3,218,216
8	ADA	Cardano	2,450,912
9	XMR	Monero	1,788,084
10	TRX	TRON	1,624,929
11	BNB	Binance Coin	1,461,507
12	MIOTA	Iota	1,448,470
13	DASH	Dash	1,368,564
14	NEO	Neo	1,108,333

References

A. F. Bariviera, M. J. Basgall, W. Hasperué, and M. Naiouf. Some stylized facts of the bitcoin market. *Physica A: Statistical Mechanics and its Applications*, 484: 82–90, 2017. ISSN 0378-4371. doi: 10.1016/j.physa.2017.04.159. URL http://www.sciencedirect.com/science/article/pii/S0378437117304697.

D. J. Bartholomew. *Analysis of multivariate social science data*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. CRC Press, Boca Raton, Florida, 2nd ed. edition, 2011. ISBN 1584889616.

D. G. Baur, T. Dimpfl, and K. Kuck. Bitcoin, gold and the us dollar – a replication and extension. *Finance Research Letters*, 25:103–110, 2018. ISSN 1544-6123. doi: 10.1016/j.frl.2017.10.012. URL http://www.sciencedirect.com/science/article/pii/S1544612317305093.

C. T. Bergstrom and L. A. Dugatkin. *Evolution*. W.W. Norton & Company, New York, second edition edition, 2016. ISBN 0393601048.

S. Blackburn. *The Oxford dictionary of philosophy: Over 3000 entries*. Oxford reference online. Oxford Univ. Press, Oxford, 2. ed. rev edition, 2008. ISBN 0199541434.

N. Borri. Conditional tail-risk in cryptocurrency markets. *Journal of Empirical Finance*, 50:1–19, 2019. ISSN 09275398. doi: 10.1016/j.jempfin.2018.11.002.

G. M. Caporale, L. Gil-Alana, and A. Plastun. Persistence in the cryptocurrency market. *Research in International Business and Finance*, 46:141–148, 2018. ISSN 0275-5319. doi: 10.1016/j.ribaf.2018.01.002. URL http://www.sciencedirect.com/science/article/pii/S0275531917309200.

B. A. Cerny and H. F. Kaiser. A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate behavioral research*, 12(1):43–47, 1977. ISSN 0027-3171. doi: 10.1207/s15327906mbr1201{\textunderscore}3.

S. Corbet, B. Lucey, A. Urquhart, and L. Yarovaya. Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, 62:182–199, 2019. ISSN 10575219. doi: 10.1016/j.irfa.2018.09.003. URL http://www.sciencedirect.com/science/article/pii/S1057521918305271.

N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, Cambridge, 2000. ISBN 9780511801389. doi: 10.1017/CBO9780511801389.

A. H. Dyhrberg. Bitcoin, gold and the dollar – a garch volatility analysis. *Finance Research Letters*, 16:85–92, 2016a. ISSN 1544-6123. doi: 10.1016/j.frl.2015.10.008. URL http://www.sciencedirect.com/science/article/pii/S1544612315001038.

A. H. Dyhrberg. Hedging capabilities of bitcoin. is it the virtual gold? *Finance Research Letters*, 16:139–144, 2016b. ISSN 1544-6123. doi: 10.1016/j.frl.2015.10.025.

A. ElBahrawy, L. Alessandretti, A. Kandler, R. Pastor-Satorras, and A. Baronchelli. Evolutionary dynamics of the cryptocurrency market. *Royal Society open science*, 4(11):170623, 2017. ISSN 2054-5703. doi: 10.1098/rsos.170623.

W. Feng, Y. Wang, and Z. Zhang. Can cryptocurrencies be a safe haven: a tail risk perspective analysis. *Applied Economics*, 50(44):4745–4762, 2018. ISSN 0003-6846. doi: 10.1080/00036846.2018.1466993.

R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936. ISSN 20501420. doi: 10.1111/j.1469-1809.1936.tb02137.x.

W. Härdle and L. Simar. *Applied multivariate statistical analysis: [R & Matlab codes]*. Springer, Berlin and Heidelberg, 3. ed. edition, 2012. ISBN 3642172296.

W. K. Härdle, C. Harvey, and R. Reule. Understanding cryptocurrencies. *IRTG 1792 Discussion Paper*, 2018-044, 2018. URL https://www.wiwi.hu-berlin.de/de/forschung/irtg/results/discussion-papers/discussion-papers-2017-1/irtg1792dp2018-044.pdf.

I. Henriques and P. Sadorsky. Can bitcoin replace gold in an investment portfolio? *Journal of Risk and Financial Management*, 11(3):48, 2018. doi: 10.3390/jrfm11030048.

D. W. Hosmer and S. Lemeshow. *Applied logistic regression*. A Wiley-Interscience publication. John Wiley, New York, 2nd ed., [repr.] edition, 2010. ISBN 0471356328.

H. E. Hurst. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:770–808, 1951.

J. Istas and G. Lang. Quadratic variations and estimation of the local hölder index of a gaussian process. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 33(4):407–436, 1997. ISSN 02460203. doi: 10.1016/S0246-0203(97)80099-4.

Y. Jiang, H. Nie, and W. Ruan. Time-varying long-term memory in bitcoin market. *Finance Research Letters*, 25:280–284, 2018. ISSN 1544-6123. doi: 10.1016/j.frl.2017.12.009.

G. Julián-Moreno, J. E. López de Vergara, I. González, L. de Pedro, J. Royuela-del Val, and F. Simmross-Wattenberg. Fast parallel α-stable distribution function evaluation and parameter estimation using opencl in gpgpus. *Statistics and Computing*, 27(5):1365–1382, 2017. ISSN 0960-3174. doi: 10.1007/s11222-016-9691-9.

H. F. Kaiser. An index of factorial simplicity. *Psychometrika*, 39(1):31–36, 1974. ISSN 0033-3123. doi: 10.1007/BF02291575.

H. F. Kaiser. A revised measure of sampling adequacy for factor-analytic data matrices. *Educational and Psychological Measurement*, 41(2):379–381, 1981. ISSN 0013-1644. doi: 10.1177/001316448104100216.

T. Klein, H. Pham Thu, and T. Walther. Bitcoin is not the new gold – a comparison of volatility, correlation, and portfolio performance. *International Review of Financial Analysis*, 59:105–116, 2018. ISSN 10575219. doi: 10.1016/j.irfa.2018.07.010.

I. A. Koutrouvelis. Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75(372):918, 1980. ISSN 01621459. doi: 10.2307/2287182.

I. A. Koutrouvelis. An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics - Simulation and Computation*, 10 (1):17–28, 1981. ISSN 0361-0918. doi: 10.1080/03610918108812189.

J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif., 1967. University of California Press. URL https://projecteuclid.org/euclid.bsmsp/1200512992.

M. Mahner and M. Kary. What exactly are genomes, genotypes and phenotypes? and what about phenomes? *Journal of theoretical biology*, 186(1):55–63, 1997. ISSN 0022-5193. doi: 10.1006/jtbi.1996.0335.

N. J. Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991. ISSN 0006-3444. doi: 10.1093/biomet/78.3.691.

W. T. Parry and E. A. Hacker. *Aristotelian logic*. State University of New York Press, Albany, 1991. ISBN 9780791406908.

S. T. Rachev and S. Mittnik. *Stable Paretian models in finance*. Series in financial economics and quantitative analysis. J. Wiley, Chichester, 2000. ISBN 9780471953142.

L. H. Rieseberg, S. A. Church, and C. L. Morjan. Integration of populations and differentiation of species. *The New phytologist*, 161(1):59–69, 2004. doi: 10.1046/j.1469-8137.2003.00933.x.

R. Selmi, W. Mensi, S. Hammoudeh, and J. Bouoiyour. Is bitcoin a hedge, a safe haven or a diversifier for oil price movements? a comparison with gold. *Energy Economics*, 74:787–801, 2018. ISSN 01409883. doi: 10.1016/j.eneco.2018.07.007.

D. Stosic, D. Stosic, T. B. Ludermir, and T. Stosic. Collective behavior of cryptocurrency price changes. *Physica A: Statistical Mechanics and its Applications*, 507:499–509, 2018. ISSN 0378-4371. doi: 10.1016/j.physa.2018.05.050.

B. G. Tabachnick and L. S. Fidell. *Using multivariate statistics*. Pearson Education, Boston, 6th ed. edition, 2013. ISBN 0205890814.

T. Takaishi. Statistical properties and multifractality of bitcoin. *Physica A: Statistical Mechanics and its Applications*, 506:507–519, 2018. ISSN 0378-4371. doi: 10.1016/j.physa.2018.04.046.

Transparency Market Research. Cryptocurrency market to reach US$6,702.1 mn by 2025, 2018. URL https://www.transparencymarketresearch.com/cryptocurrency-market.html.

A. Urquhart. The inefficiency of bitcoin. *Economics Letters*, 148:80–82, 2016. ISSN 01651765. doi: 10.1016/j.econlet.2016.09.019.

J. D. Washburn, K. A. Bird, G. C. Conant, and J. C. Pires. Convergent evolution and the origin of complex phenotypes in the age of systems biology. *International Journal of Plant Sciences*, 177(4):305–318, 2016. doi: 10.1086/686009.

W. C. Wei. Liquidity and market efficiency in cryptocurrencies. *Economics Letters*, 168:21–24, 2018. ISSN 01651765. doi: 10.1016/j.econlet.2018.04.003.

Y. G. Yatracos. Variance and clustering. *Proceedings of the American Mathematical Society*, 126(04):1177–1179, 1998. ISSN 00029939. doi: 10.1090/S0002-9939-98-04524-9.

Y. G. Yatracos. The asymptotic distribution of a cluster-index for i.i.d. normal random variables. *The Annals of Applied Probability*, 19(2):585–595, 2009. ISSN 1050-5164. doi: 10.1214/08-AAP553.

Y. G. Yatracos. Detecting clusters in the data from variance decompositions of its projections. *Journal of Classification*, 30(1):30–55, 2013. ISSN 0176-4268. doi: 10.1007/s00357-013-9124-9.

W. Zhang, P. Wang, X. Li, and D. Shen. Some stylized facts of the cryptocurrency market. *Applied Economics*, 50(55):5950–5965, 2018. ISSN 0003-6846. doi: 10.1080/00036846.2018.1488076.

IRTG 1792 Discussion Paper Series 2019

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Cooling Measures and Housing Wealth: Evidence from Singapore" by Wolfgang Karl Härdle, Rainer Schulz, Taojun Xie, January 2019.

002 "Information Arrival, News Sentiment, Volatilities and Jumps of Intraday Returns" by Ya Qian, Jun Tu, Wolfgang Karl Härdle, January 2019.

003 "Estimating low sampling frequency risk measure by high-frequency data" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

004 "Constrained Kelly portfolios under alpha-stable laws" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

005 "Usage Continuance in Software-as-a-Service" by Elias Baumann, Jana Kern, Stefan Lessmann, February 2019.

006 "Adaptive Nonparametric Community Detection" by Larisa Adamyan, Kirill Efimov, Vladimir Spokoiny, February 2019.

007 "Localizing Multivariate CAViaR" by Yegor Klochkov, Wolfgang K. Härdle, Xiu Xu, March 2019.

008 "Forex Exchange Rate Forecasting Using Deep Recurrent Neural Networks" by Alexander J. Dautel, Wolfgang K. Härdle, Stefan Lessmann, Hsin-Vonn Seow, March 2019.

009 "Dynamic Network Perspective of Cryptocurrencies" by Li Guo, Yubo Tao, Wolfgang K. Härdle, April 2019.

010 "Understanding the Role of Housing in Inequality and Social Mobility" by Yang Tang, Xinwen Ni, April 2019.

011 "The role of medical expenses in the saving decision of elderly: a life cycle model" by Xinwen Ni, April 2019.

012 "Voting for Health Insurance Policy: the U.S. versus Europe" by Xinwen Ni, April 2019.

013 "Inference of Break-Points in High-Dimensional Time Series" by Likai Chen, Weining Wang, Wei Biao Wu, May 2019.

014 "Forecasting in Blockchain-based Local Energy Markets" by Michael Kostmann, Wolfgang K. Härdle, June 2019.

015 "Media-expressed tone, Option Characteristics, and Stock Return Predictability" by Cathy Yi-Hsuan Chen, Matthias R. Fengler, Wolfgang K. Härdle, Yanchu Liu, June 2019.

016 "What makes cryptocurrencies special? Investor sentiment and return predictability during the bubble" by Cathy Yi-Hsuan Chen, Roméo Després, Li Guo, Thomas Renault, June 2019.

017 "Portmanteau Test and Simultaneous Inference for Serial Covariances" by Han Xiao, Wei Biao Wu, July 2019.

018 "Phenotypic convergence of cryptocurrencies" by Daniel Traian Pele, Niels Wesselhöfft, Wolfgang K. Härdle, Michalis Kolossiatis, Yannis Yatracos, July 2019.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.