Phenotypic convergence of cryptocurrencies

Daniel Traian Pele[a], Niels Wesselhöfft[b], Wolfgang K. Härdle[c]

[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, Spandauer Str. 1, 10178 Berlin, Germany, E-mail: wesselhn@hu-berlin.de*

[c]*Humboldt-Universität zu Berlin, IRTG 1792, Spandauer Str. 1, 10178 Berlin, Germany and School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899*

Abstract

In this paper we show that the financial asset class of cryptocurrencies converges to a unique asset class in comparison to stocks, commodities and exchange rates. In biological terms, individual cryptocurrencies respond to similar selection pressures by developing similar characteristics, leading to a phenotypic convergence of cryptocurrencies. In order to retrieve the phenotype of cryptocurrencies, we find the proximal genus and the specific difference (*genus proximum et differentia specifica*) for the daily time series of cryptocurrencies returns, comparing them to classical asset returns. In this sense, the daily time series of asset returns 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, Quadratic Discriminant Analysis, Support Vector Machines) we are able to classify cryptocurrencies as a new asset class with unique features in the tails of the returns distribution. By using expanding window estimation for the factors, we observe a divergent evolution of the cryptocurrencies class, mainly due to the heavy tail behaviour of the respective return distributions. The codes utilized are available via www.quantlet.de.

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

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 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 (BTC) can be found in Corbet et al. (2018). Table 1 lists the 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 validity of phenotypic convergence in case of cryptocurrencies. 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 price or returns. In order to derive the assets phenotype, we are using a genus-differentia approach, allowing to separate the cryptocurrencies universe from the classical assets.

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 find the proximal genus and the specific difference (*genus proximum et differentia specifica*) of the daily time series of cryptocurrencies returns.

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 USD.
Dyhrberg (2016b)	BTC, Federal funds rate, USD/EUR, USD/GBP, FTSE index, Gold futures, Gold cash.	2010-2015, daily data.	BTC ranges in between a USD and a 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). Standard deviation of BTC 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 degree of persistence 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 West Texas Intermediate (WTI), S&P500 index, MSCI World and MSCI Emerging Markets 50 index.	2011-2017, daily data.	BTC returns have the highest mean and standard deviation among all assets.
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.

Trough the means of dimensionality reduction techniques (like Factor Analysis) and classification techniques (like Binary Logistic Regression and Support Vector Machines), 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. 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 most important result of our paper is the discovery of a phenotypic convergence of cryptocurrencies, compared to the classical assets. 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 first section describes the statistical methodology used, including Factor analysis, Logistic regression, Support vector machines (SVM) and the evolutive divergence. The second section describes the data-set and interprets the results of the classification; the third 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.

1. 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 – FA,) in order to retain the orthogonal factors which maximizes the explained variance. Third, we employ classification techniques (Binary logistic regression, Discriminant analysis, Support Vector Machine – SVM) 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 poses specific characteristics allowing them to differentiate over time from the classical assets.

1.1. 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 define the following notations:

- 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 $i = 1 \ldots n$, $t = 1 \ldots T$;
- $R = (R_{i,t})^{\mathrm{T}}_{i=1\ldots,T} \in M(T, n)$ - 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 $X_t = (x_{it_t,k})_{\substack{i=1\ldots n \\ k=1\ldots p}} \in M(n, p)$, whose components are detailed below, estimated for the time interval [1, t], with $t = t_0 \ldots T$, where $t_0 = [T/3]$.
First, we took into account the 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, in order to capture some scaling properties:

- the tail parameter: $Stable\_\alpha_{it}$;
- the scale parameter: $Stable\_\gamma_{it}$.

The α-stable parameters were estimated using the empirical characteristic function method, following Koutrouvelis (1980) and Koutrouvelis (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 was 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 $\mathscr{X} \in \mathrm{R}^{n \times p \times T'}$, 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.

1.2. 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 are using Factor Analysis to extract the main factors explaining the variation in the initial dataset, the reason for this being the fact that PCA is a linear combination of variables, while Factor Analysis is a measurement model of a latent variable. The aim of the factor analysis (FA) 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:

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

$$(p \times 1) = (p \times k)(k \times 1) + (p \times 1) + (p \times 1), \tag{2}$$

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.
- 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 was applied on the entire dataset X_T, the p initial variables being estimated for the entire time period $[1, \dots, T]$. The p-dimensional dataset X_T was then projected on the k-dimensional space defined by the k orthogonal factors, in order to observe a separation of the assets.

1.3. Assets Classification

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

1.3.1. 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:

$$P(Y_i = 1) = \frac{\exp(\beta_{0j} + \beta_{1j} F_{ji})}{1 + \exp(\beta_{0k} + \beta_{1j} F_{ji})}, \tag{3}$$

where $Y_i = 1$ for cryptocurrencies, $Y_i = 0$ for classical assets, and $F_j, j \in \{1, \dots, k\}$ are the k orthogonal factors retrieved through the factor analysis. Based on the explanatory power and the significance of the model 3, we can derive the most important factors contributing to the *differentia specifica* of cryptocurrencies. As a performance measure for the model 3, we are using $\tilde{R}^2$ (Nagelkerke, 1991), where:

$$\tilde{R}^2 = \frac{1 - \left\{ \frac{L(\mathbf{0})}{L(\widehat{\boldsymbol{\beta}})} \right\}^{\frac{2}{n}}}{1 - \{L(\mathbf{0})\}^{\frac{2}{n}}}. \tag{4}$$

In formula 4, $L(0)$ is the likelihood of the intercept-only model, while $L(\widehat{\boldsymbol{\beta}})$ is the likelihood of the full model.

1.3.2. 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: $X_i \sim N(\mu_i, \Sigma_i)$, where X_i belong to class ω_i, $\Sigma_i = \Sigma_j$. The goal is to project samples X onto a line $Z = w^\top X$, where we select the projection that maximizes the separability. Specifically, we maximize the normalized, squared distance in the means of the classes

$$w^* = \arg\max_w \frac{|w^\top(\mu_i - \mu_j)|^2}{s_i^2 + s_j^2}, \tag{5}$$

$$s_i^2 = \sum_{x_i \in \omega_i} (w^\top x_i - w^\top \mu_i)^2 = w^\top S_i w, \tag{6}$$

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

$$w^* = S_W^{-1}(\mu_i - \mu_j), \; S_W = S_i + S_j. \tag{7}$$

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 $\Sigma_i \neq \Sigma_j$, allowing for a non-linear classifier.

1.3.3. Support Vector Machines

Support Vector Machines (SVM) are a data classification technique, aiming 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 $D = (X_1, Y_1), \dots (X_n, Y_n)$, $X_i \in \mathbb{R}^2, \quad Y_i \in [0,1]$, the aim is to a hyperplane that maximizes the margin

$$\min_{w,b} \frac{1}{2}\|w\|^2, \text{ s.t. } \quad Y_i\left(w^\top X_i + b\right) \geq 1, i = 1, \dots, n. \tag{8}$$

1.3.4. 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 $t = t_0, \dots, T$, where $t_0 = [T/3]$, the p-dimensional dataset X_t is projected on the k-dimensional space defined by the main factors extracted trough 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.

2. 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, the total market capitalization being less than 16 billion US dollars (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/`, for the time interval 10/20/2014 - 10/16/2018. The CRyptocurrency IndeX is a benchmark for the cryptocurrencies market, being realtime computed by the Ladislaus von Bortkiewicz Chair of Statistics at Humboldt University Berlin, Germany. The 15 components of the CRIX index (see the Appendix A) account for 90% of the total cryptocurrencies market capitalization, as seen in

[Figure 1](). In our analysis, the USDT cryptocurrency was discarded due to the fact that it is an atypical digital currency, having very little variation around the reference value of 1 USD.

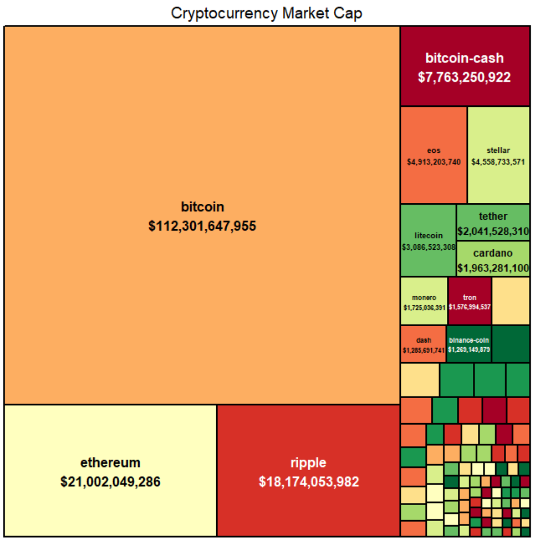

Figure 1: Cryptocurrencies market capitalization (USD) at 10/19/2018. Mkt_cryptos.

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

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

2.1. 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.

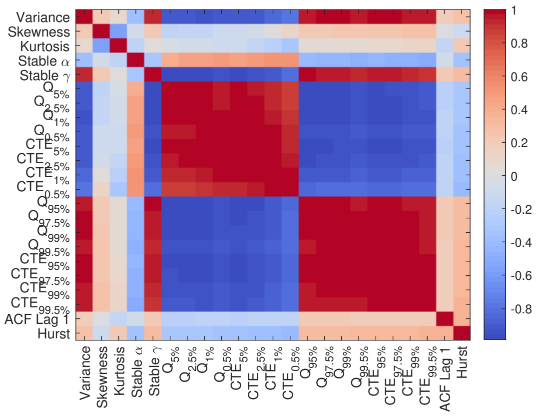

Figure 2: Correlation matrix. SFA_cryptos

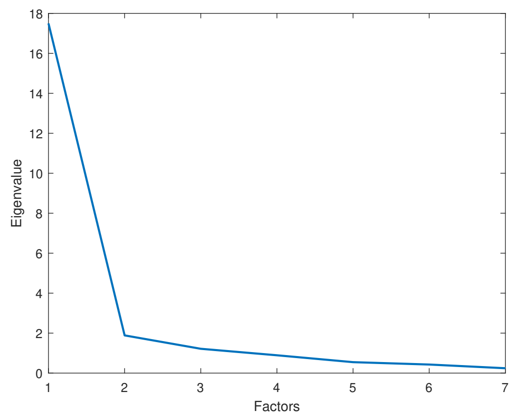

Figure 3: Scree plot. SFA_cryptos

Based on the eigenvalues criteria, we can select those factors for which the eigenvalue is higher than 1 (see the 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); the KMO test should be greater than 0.5 for a satisfactory factor analysis (Tabachnick and Fidell, op. 2013).

The overall KMO test is computed using the following formula:

$$KMO = \frac{\sum_i \sum_{i \neq j} r_{ij}^2}{\sum_i \sum_{i \neq j} r_{ij}^2 + \sum_i \sum_{i \neq j} u_{ij}^2} \tag{9}$$

where $R = r_{ij}$ is the correlation matrix and $U = u_{ij}$ is the partial covariance matrix (Cerny and Kaiser, 1977, Kaiser, 1974).

The individual KMO test is computed using the formula:

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} u_{ij}^2} \tag{10}$$

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

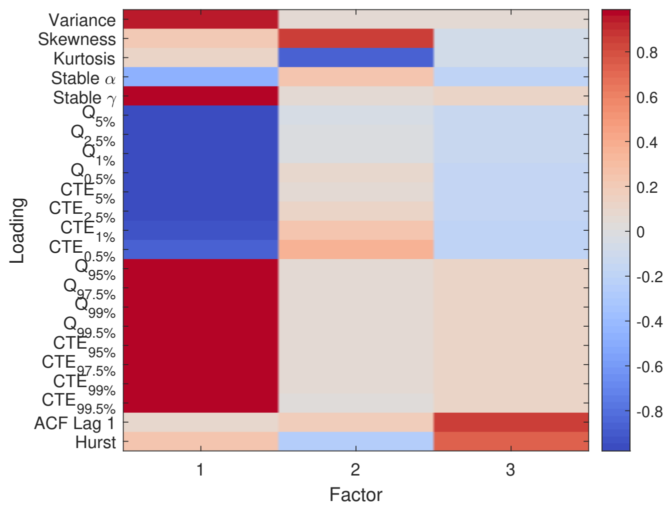

Figure 4: Loadings of the three factors. SFA_cryptos.

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

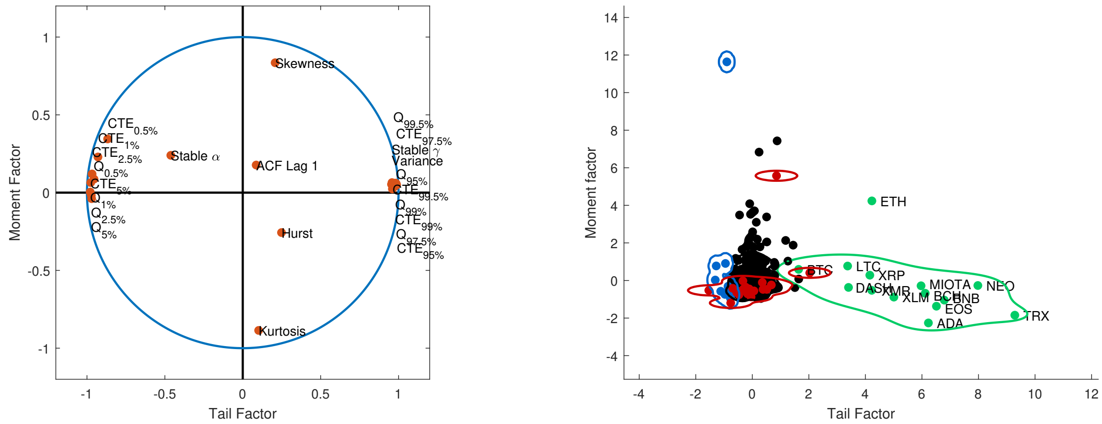

Figure 5: Loadings (left) and scores (right) based on tail and moment factor. SFA_cryptos.

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.

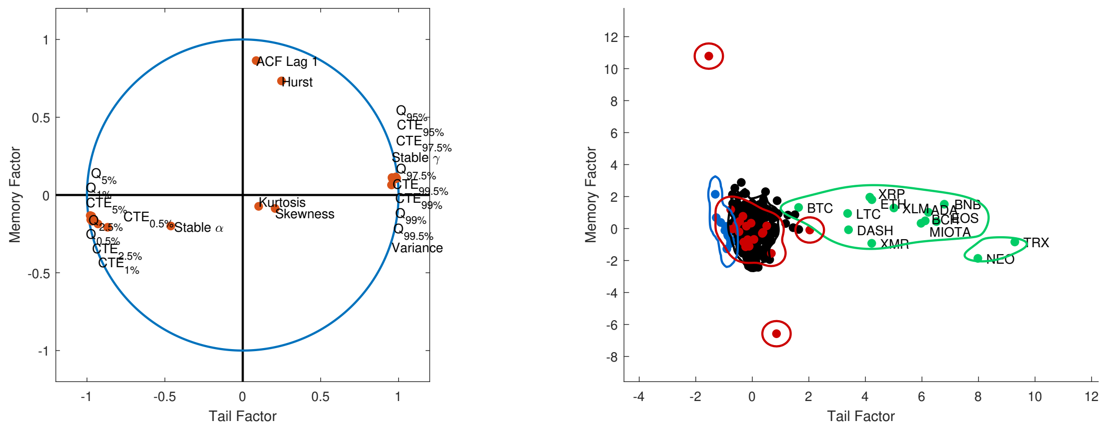

Figure 6: Loadings (left) and scores (right) based on tail and memory factor. SFA_cryptos.

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 subliminal importance (see Figure 7).

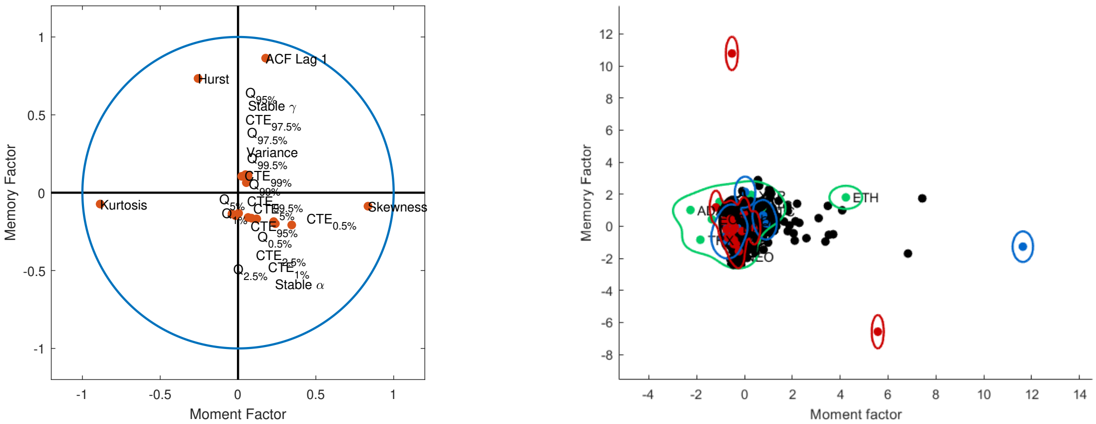

Figure 7: Loadings (left) and scores (right) based on moment and memory factor. SFA_cryptos.

Based on the data revealed in the 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. it's 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	**Estimate**	**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

2.2. *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:

$$P(Y_i = 1) = \frac{\exp(\beta_{0j} + \beta_{1j} F_{ji})}{1 + \exp(\beta_{0k} + \beta_{1j} F_{ji})}, \tag{11}$$

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

Table 5 lists the estimated β_{1j} of the binary logistic regression model 11, with the performance measure defined by the equation 4.

Table 5: Estimates of the model 11

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

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

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.
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 the Table 4 for Bitcoin's profile).

The same conclusion can be drawn by looking at the results of the Support

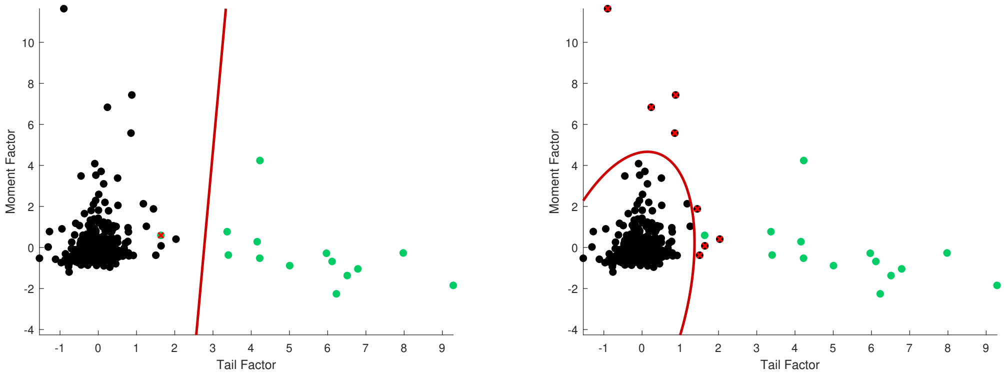

Figure 8: Discriminant Analysis: linear(left) and quadratic (right). Green dots denote the cryptocurrencies, while the black dots denote the other assets; the dots highlighted in red are cases of misclassification. SFA_cryptos.

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).

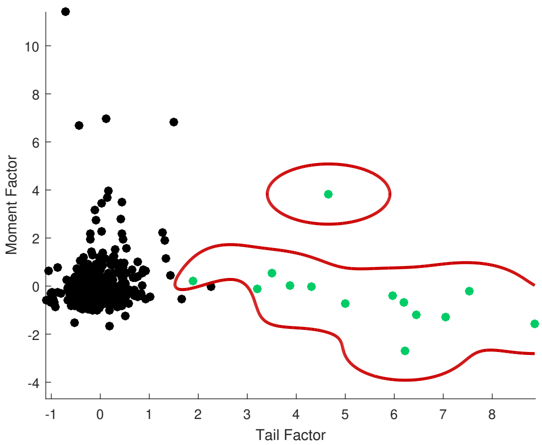

Figure 9: Support Vector Machines. Green dots denote the cryptocurrencies, while the black dots denote the other assets. SFA_cryptos.

From an Aristotelian point of view, we can conclude that the differentia specifica of the cryptocurrencies 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 alpha stable scale parameter and value of the alpha stable tail index closer to 1.

3. 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 $t = t_0, \dots, T$, the p-dimensional dataset is projected on the k-dimensional space defined by the main factors extracted trough 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. The 23-dimension dataset is estimated first for the time interval $[1, t_0]$=[10/10/2014, 02/19/2016]. Then, the time window is extended on a daily basis, up to T=10/16/2018 and for each step in time, the 23-dimension dataset is projected on the 2-dimension space defined by the tail factor and the moment factor, estimated for the entire time period.

Figure 10 presents a snapshot of the evolution of the assets universe using the expanding window approach.

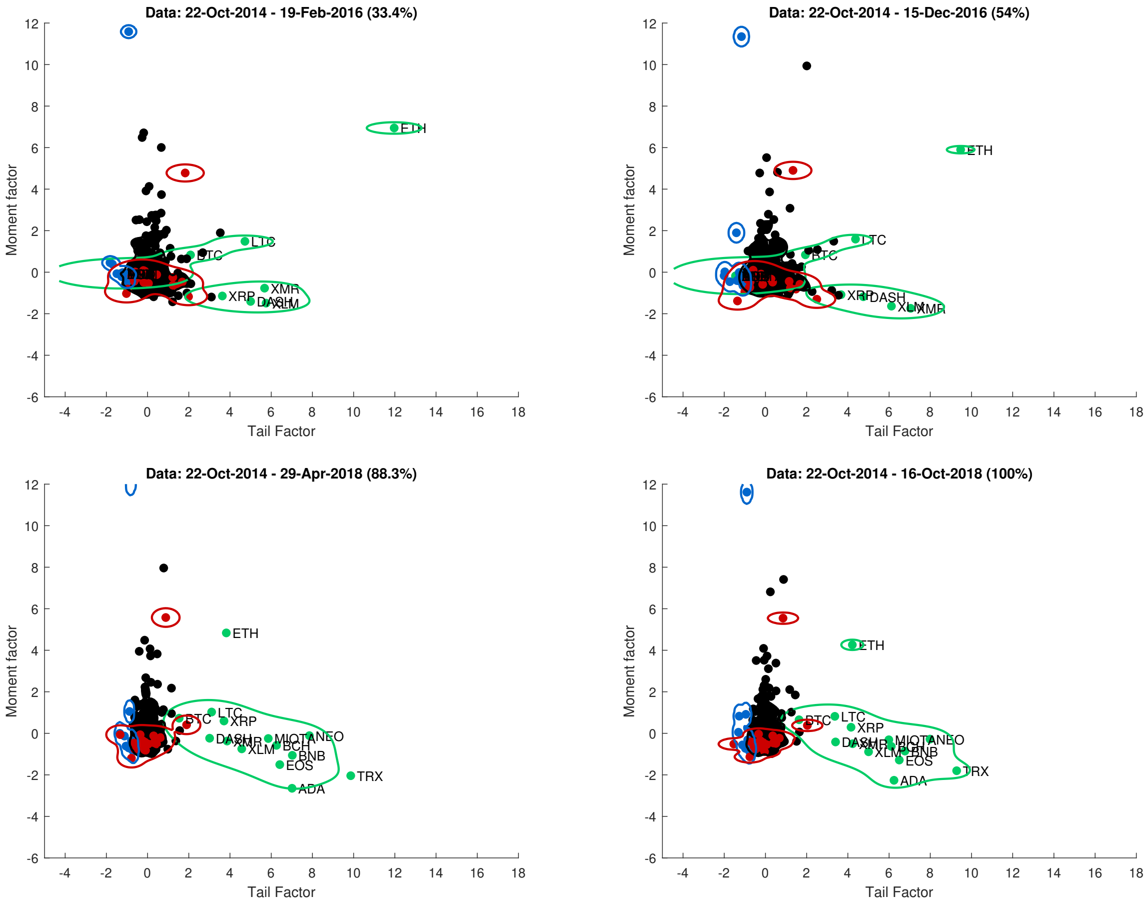

Figure 10: The evolution of the assets universe using the expanding window approach. The colour code is the following: green - cryptocurrencies, black – stocks, red – commodities, blue – exchange rates. DFA_cryptos.

The daily evolution of the assets universe, for the period 02/19/2016-10/16/2018 is depicted in this movie.

Looking at the evolution of the assets universe, it appears that the behaviour of cryptocurrencies is a dialectical one and 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 certain characteristics (phenotypic convergence) that make them fully distinguishable from the classical assets (divergent evolution).

In order to test this behaviour, we are using the Likelihood Ratio associated to model 3, estimated using the expanding window approach previously described. Thus, the Likelihood Ratio of the model 3 can be defined as:

$$LR(\widehat{\boldsymbol{\beta}}) = -2(\log L(\widehat{\boldsymbol{\beta}}) - \log L(\widehat{\boldsymbol{\beta}_s})), \tag{12}$$

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 equation 12 is called deviance (Hosmer and Lemeshow, 2010) and is a measure of model goodness-of-fit, 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:

$$D = [LR(\widehat{\boldsymbol{\beta}}) - LR(0)] \sim \chi^2(1), \tag{13}$$

$LR(0)$ 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 11) and the p-value associated to equation 13 (see the Figure 12).

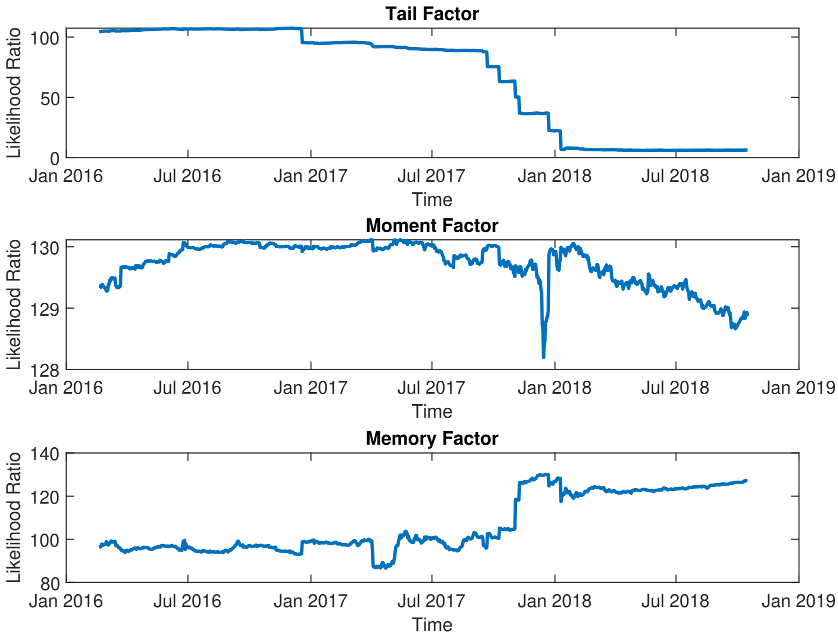

Figure 11: Likelihood Ratios for the model 3, estimated for the period 02/19/2016-10/16/2018. CONV_cryptos.

As shown in Figure 11, the classification power of the moment and memory factor is negligible, as the Likelihood Ratio is significantly higher than 0. However, the model based on the tail factor has an improving goodness-of-it over time, the Likelihood Ratio converging to lower values regime. This result is augmented by the evolution of P-values, shown in Figure 12: the tail factor is the only one able to discriminate between the cryptocurrencies and classical assets.

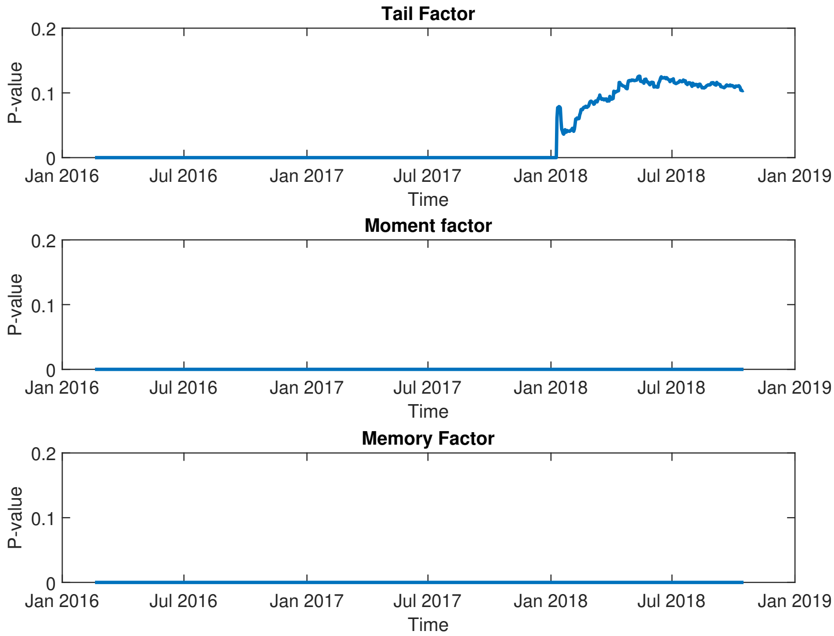

Figure 12: P-values for the equation 13, estimated for the period 02/19/2016-10/16/2018. CONV_cryptos.

By looking the 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.

4. Conclusions

In this paper we applied linear Factor models on statistics of log returns in order to discriminate between the cryptocurrencies and traditional assets: stocks, exchange rates and commodities. Utilizing a multidimensional approach, which takes various indicators into account, which describe the market risk behaviour, tail behaviour and long-memory characteristics of the time series of daily log-returns, we found the proximal genus and the specific difference (genus proximum et differentia specifica) between the daily time series of cryptocurrencies returns and the classical assets returns.

Through the means of dimensionality reduction techniques and classification techniques, we showed that large parts 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 moment factor and the memory factor 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.

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.

Viewing 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

We would like to thank the editor and the anonymous referees for their valuable comments to this article. 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 assets 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	USDT	Tether	2,755,619
9	ADA	Cardano	2,450,912
10	XMR	Monero	1,788,084
11	TRX	TRON	1,624,929
12	BNB	Binance Coin	1,461,507
13	MIOTA	Iota	1,448,470
14	DASH	Dash	1,368,564
15	NEO	Neo	1,108,333

The components of the S&P500 used in the analysis and the entire list of assets can be found here.

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.

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. M. Lucey, A. Urquhart, and L. Yarovaya. Cryptocurrencies as a financial asset: A systematic analysis. *SSRN Electronic Journal*, 2018. doi: 10.2139/ssrn.3143122.

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.

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.

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, op. 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.

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.