Are Cryptos becoming alternative Assets?

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

[a]*Corresponding author, 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*
[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
Department of Information Management and Finance, National Chiao Tung University, Taiwan, ROC
[d]*Department of Mathematics and Statistics, University of Cyprus, Nicosia, 1678, Cyprus*
Central Bank of Cyprus, Nicosia 1076, Cyprus[1]
[e]*Yau Mathematical Sciences Center, Tsinghua University, Beijing, China*

Abstract

This research provides insights for the separation of cryptocurrencies from other assets. Using dimensionality reduction techniques, we show that most of the variation among cryptocurrencies, stocks, exchange rates and commodities can be explained by tail, moment and memory factors of their log-returns. By applying various classification models, we are able to classify cryptocurrencies as a separate asset class, mainly due to the tail factor. The main result is the complete separation of cryptocurrencies from the other asset types, using the Maximum Variance Components Split method. Additionally, we show that cryptocurrencies tend to exhibit similar characteristics over time and become more distinguished from other asset classes (synchronic evolution).

Keywords: cryptocurrency, classification, multivariate analysis, factor models, variance components split methods, synchronic evolution
JEL Classification: C1, G1

1. Introduction

Cryptocurrencies, seen as new digital currencies, have attracted much attention from investors and academics. Searching for "cryptocurrencies" on Google Scholar returns more than 30,000 items, as of June 30, 2020. Most research articles focus on Bitcoin (BTC), as it is considered the first cryptocurrency and has the largest capitalization since its inception; see for example, Dyhrberg (2016a), Bariviera et al. (2017). More

[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.

recently, an extensive literature review on Bitcoin can be found in Corbet et al. (2019). Along with this growing popularity, the market capitalization of cryptocurrencies was increasing substantially; the total capitalization for cryptocurrencies market was around US$ 266 billion as of June 20, 2020, from almost US$ 9 billion as of June 20, 2014 (https://coinmarketcap.com).

Despite their growing popularity, there is no widely accepted definition of cryptocurrencies which would allow one to identify them within the existing economic theory (Núñez et al., 2019). In general, cryptocurrencies are defined as "digital representations of value, made possible by advances in cryptography and distributed ledger technology (DLT)" (I.M.F., 2019).

Since it appears to be difficult to reach consensus on a standard definition of cryptocurrencies, one can define cryptocurrencies by investigating whether their returns behave similarly to other asset classes (Liu and Tsyvinski, 2018). One of the approaches used to demonstrate this is by analyzing the properties of log-returns distribution. Most of the research shows that cryptocurrencies present long-range memory (Bariviera et al. (2017), Caporale et al. (2018), Jiang and Han (2018)), multifractality (Takaishi, 2018) higher volatility, skewness and kurtosis compared to classical assets (Härdle et al. (2018), Klein et al. (2018)). Borri (2019) shows that cryptocurrencies exhibit large and volatile return swings and are riskier than most of the other assets, while Zhang et al. (2018) find that cryptocurrencies exhibit heavy tails, quickly decaying returns autocorrelations, slowly decaying autocorrelations for absolute returns, strong volatility clustering, leverage effects, long-range dependence and power-law correlation between price and volume.

Another approach used to separate cryptocurrencies from classical assets is to develop models able to explain specificities of cryptocurrencies. For example, Manavi et al. (2020), using the matrix correlation method, compare 7 cryptocurrencies with a sample of the three types of monetary systems: 28 exchange rates, 2 commodities, 2 commodity-based indices, and 3 financial market indices. Their results show that the cryptocurrency market and Forex market belong to different system communities.

Using a different approach, Liu and Tsyvinski (2018) analyse the relationship between cryptocurrencies and each of the following asset classes: stocks, precious metals and currencies. They show that the risk-return tradeoff of three major cryptocurrencies (Bitcoin, Ripple, and Ethereum) is distinct from those of the above asset categories. They also show that the cryptocurrency returns can be predicted by factors which are specific to cryptocurrency markets – momentum and investor attention. More recently, Liu et al. (2019) developed a three-factor model using the CAPM approach (Fama and French, 1996) and showed that the cross-sectional expected cryptocurrency returns can be captured by three factors: the market factor, the size factor and momentum factor.

In line with these results, this research provides insights related to a classification of cryptocurrencies using statistical tools. Unlike the three-factor model of Liu et al. (2019), we are using dimensionality reduction techniques (Factor Analysis) applied on a dataset of risk indicators related to the empirical distribution of daily log-returns (following Bariviera et al. (2017), Caporale et al. (2018), Jiang and Han (2018)), Takaishi (2018), Härdle et al. (2018), Klein et al. (2018), Borri (2019), Zhang et al. (2018)). Our results are also based on a much larger data set, in terms of number of assets used, than in Liu and Tsyvinski (2018); we are using 150 cryptocurrencies, 496 stocks, 13 exchange rates and 20 commodities, while Liu and Tsyvinski (2018) uses only 3 cryptocurrencies (Bitcoin, Ripple, and Ethereum), 354 industries in the US, 137 industries in China, 5 exchange rates and 3 commodities.

One important result of our research is that most of the variation among cryptocurrencies, stocks, exchange rates and commodities can be explained by three factors: the tail factor, the moment factor and the memory factor. These factors are different from the ones obtained in Liu et al. (2019) and allow us to validate the complete separation of cryptocurrencies from classical assets, with their statistical characteristics forming a disjoint cluster.

Another important result of our research is the confirmation of synchronic evolution of cryptocurrencies, compared to classical assets types. Synchronic evolution refers to the fact that individual cryptocurrencies tend to develop certain similar characteristics over time that make them fully distinguishable from classical assets (ElBahrawy et al., 2017), i.e. they tend to behave like a homogeneous group, with certain statistical characteristics that individualize them in the assets ecosystem. We are able to show that cryptocurrencies have a synchronic evolution in relationship to classical assets and this evolution is driven mainly by the tail behaviour of the log-returns distribution.

Our results help build a series of recommendations based on which researchers and practitioners can differentiate cryptocurrencies from classical assets, using one of the methods presented here. The methods used include classification techniques such as Binary Logistic Regression, Discriminant Analysis, Support Vector Machines, K-means clustering and Variance Components Split (VCS). Our findings are cross-checked and validated using classical Factor Analysis, performed on a static basis and also by using an expanding window approach, where the assets universe can be observed in its evolutionary dynamic. The expanding window approach was shown to be reliable for such validations, when referring to evolutionary systems (Borri (2019), Enoksen et al. (2020)).

This research is subsequently organized as follows: Section 2 describes the methodology used; Section 3 describes the datasets and interprets the results of the classification; Section 4 describes the synchronic evolution of cryptocurrencies, while Section 5 concludes. The data and codes used to obtain the results in this paper are available via https://github.com/QuantLet/Genus_proximum_cryptos.

2. Methodology

The methodology used in this paper has four layers: Layer 1, where we describe the multidimensional dataset used to assess the behaviour of time series of assets' daily log-returns; Layer 2, where we apply data dimension reduction and orthogonalization methods (Factor Analysis) on the dataset described in Layer 1, in order to retain the orthogonal factors which maximize the explained variance and could discriminate between cryptocurrencies and classical assets; Layer 3, where we use classification techniques (Binary Logistic Regression, Discriminant analysis, Support Vector Machine, K-means clustering, Variance Components Split methods) to separate cryptocurrencies from classical assets, using the factors estimated in Layer 2; Layer 4, where we confirm the validity of the synchronic evolution property of cryptocurrencies showing their specific characteristics that differentiate over time from classical assets, using the projection of the multidimensional dataset described in Layer 1 on the 3D space defined by the factors extracted in Layer 2.

2.1. Layer 1 - Multidimensional dataset

The initial dataset consist of daily log-returns of the assets. In order to properly classify the assets within the assets universe, we need a dataset of variables-indicators that have the statistical power to differentiate between cryptocurrencies and classical assets (stocks, exchange rates and commodities). These indicators are estimates of model paramaters associated with the daily log-returns. We denote by n the number of assets in the dataset, by t the time index, $t \in \{1, \ldots, T\}$, where T is the time of the last record in the dataset and by $p = 23$ is the number of indicators. The daily log-return for asset i in day t, is denoted as $R_{i,t} = \log P_{i,t} - \log P_{i,t-1}$, with $i = 1 \ldots n$, $t = 1 \ldots T$, where $P_{i,t}$ is the closing price for asset i in day t. The dataset of indicators can be seen as a tensor $\mathscr{X} \in \mathbb{R}^{n \times p \times T'}$, where $T' = T - t_0$ is the number of time points. The components of the matrix $\mathscr{X}_t = (x_{it,j})_{\substack{i=1\ldots n \\ j=1\ldots p}} \in \mathbb{R}^{n \times p}$, detailed below, are estimates for the time interval $[1, t]$, with $t = t_0, .., T$, where $t_0 = [T/2]$ (the integer part of $T/2$).

Most of the variables-indicators used for taxonomy are selected from the indicators already validated in the literature to differentiate between cryptocurrencies and classical assets. First, we took into account the central moments of the log-returns distribution, through the following parameters: variance ($\sigma^2_{i,t}$), skewness ($\mathrm{S}_{i,t}$) and Kurtosis ($\mathrm{K}_{i,t}$) (used in Bariviera et al. (2017), Härdle et al. (2018), Takaishi (2018)). Second, we estimated the following parameters of the α-stable distribution, fitted to daily log-returns, in order to capture tail dependent behavior: the tail exponent ($Stable\_\alpha_{i,t} \in (0, 2]$, with lower values indicating heavier tails) and the scale parameter ($Stable\_\gamma_{i,t} \geq 0$). 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 (for their use in cryptocurrencies market, see Li et al.

(2019) and Schnaubelt et al. (2019)). The α-stable parameters were estimated using the empirical characteristic function method, following Koutrouvelis (1980, 1981), through the Matlab library *stbl* (Veillete, 2012). Third, we estimated the quantiles and the conditional tail expectations for the distribution of log-returns, in order to capture the tail behaviour (Trucíos et al., 2020): left-side quantiles $Q_{\alpha;it}$, right-side quantiles $Q_{1-\alpha;it}$, conditional left tail expectation $CTE_{\alpha,it} = \mathsf{E}\left[R_{it}|R_{it} < Q_{\alpha;it}\right]$ and conditional right tail expectation $CTE_{1-\alpha,it} = \mathsf{E}\left[R_{it}|R_{it} > Q_{1-\alpha;it}\right]$, for $\alpha \in \{0.005, 0.01, 0.025, 0.05\}$. 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 a GARCH(1,1) model, as in Conrad et al. (2018), in order to capture the ARCH/GARCH volatility model parameters. Thus, from the following variance equation of the GARCH(1,1) model estimated from daily log-returns:

$$\sigma_t^2 = \kappa + \theta_1 \sigma_{t-1}^2 + \omega_1 \varepsilon_{t-1}^2, \qquad (1)$$

we retain in our dataset the estimates of the GARCH parameter θ_{1it} and the ARCH parameter ω_{1it}.

2.2. Layer 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). Factor Analysis has been extensively used in cryptocurrencies modeling for classification purposes, e.g. Liu et al. (2019), who use it to develop the cryptocurrency 3-factor model: the market factor, the size factor and the momentum factor. PCA itself 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 of a data matrix using fewer variables, the so-called factors (Härdle and Simar, 2019). 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 matrix $\mathscr{X}_T$, the $p = 23$ variables-indicators being estimated for the entire time period $[1, T]$. The p-dimensional matrix $\mathscr{X}_T$ is then projected on the k-dimensional space defined by the k orthogonal factors (in our case $k = 3$), in order to observe a separation of the assets.

2.3. Layer 3 - Separating cryptocurrencies

In order to separate cryptocurrencies from classical assets, we are using several classification techniques: Binary Logistic Regression, Discriminant Analysis, Support Vector Machines, K-means clustering and Variance Components Split (technical details regarding these techniques can be found in Appendix A). Most of these techniques have been successfully applied in relation to cryptocurrencies and classical assets (for example Fischer et al. (2019) and Mirtaheri et al. (2009)).We show that Variance Component Split methods separate completely the cryptocurrencies data from the data of other assets.

2.4. Layer 4 - Synchronic evolution of cryptocurrencies

For observing the synchronic evolution of cryptocurrencies, we are using an expanding window approach, allowing to distinguish the convergence over time of cryptocurrencies. In fact, for $t \in \{t_0, .., T\}$, where $t_0 = [T/3]$, the p-dimensional matrix $\mathscr{X}_t$ is projected on the 3-dimensional space defined by the tail, moment and memory factors extracted through the Factor Analysis applied on the dataset $\mathscr{X}_T$. Looking at the evolution of the Likelihood Ratio from the Logistic Regression model defined in Layer 2, we can observe the ability of the tail factor to discriminate between cryptocurrencies and classical assets. In other words, cryptocurrencies develop over time similar tail behaviour, pointing out the validity of the synchronic evolution.

3. Data and Results

3.1. Multidimensional dataset

The initial dataset consist of daily log-returns of $n = 679$ assets (cryptocurrencies, commodities, exchange rates and stocks - see Table 1), covering the time period 02/01/2014 - 30/08/2019 (1478 trading days). The reason for choosing this time span for the analysis is that before 2015 the liquidity in the cryptocurrency market had been relatively low, their total market capitalization being less than US$16 billion (Feng et al., 2018). As described in Layer 1 of the methodology section 2.1, the multidimensional

Table 1: Assets used for analysis

Type of Asset	**Number of Assets**	**Source**
Cryptocurrencies	150	Coinmarketcap
Stocks	496	Bloomberg
Exchange rates	13	Bloomberg
Commodities	20	Bloomberg

dataset used for analysis is $\mathscr{X} \in \mathbb{R}^{n\times p\times T'}$, where $n = 679$ is the number of assets, $p = 23$ is the number of indicators, $T = 1478$ (corresponding to 30/08/2019), $t_0 = 739$ (corresponding to 31/10/2016) and $T' = T - t_0 = 739$ is the number of time points. The components of the matrix $\mathscr{X}_t = (x_{it,j})_{\substack{i=1\dots n\\ j=1\dots p}} \in \mathbb{R}^{n\times p}$, are estimates for the time interval $[1, t]$, with $t = t_0, .., T$.

For robustness purposes, only the assets with at least 500 observations were kept in the analysis.[2] The first component of the dataset contains a representative sample of 150 cryptocurrencies selected from the top 500 cryptocurrencies sourced from `https://coinmarketcap.com/`, accounting for 98% of total market capitalization, as of 30/08/2019. The second component contains a sample of the most traded commodities indexes (Table 2), the third component contains a sample of the most liquid exchange rates (Table 3), while the fourth component contains the constituents of the S&P500 Index, recorded at 30/08/2019. As cryptocurrencies daily data are available at all times, while the stocks data obtained from Bloomberg observe market closure days (weekends and public holidays), the cryptocurrency data were pre-processed and the log-returns were computed in the same way as for classical assets (for example, the Monday return compares the Monday closing price to the Friday closing price). A detailed analysis regarding data comparability and analysis of cryptocurrencies can be found in Alexander and Dakos

[2]The complete list of the assets included in the analysis can be found in the file https://github.com/QuantLet/Genus_proximum_cryptos/blob/master/list.xlsx.

(2020).

Table 2: 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 3: 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

3.2. Factor Analysis

Factor Analysis is a classical method used to find latent variables or factors among observed variables, by grouping variables with similar characteristics. For this purpose, we are using the matrix $\mathscr{X}_T = (x_{iT,j})_{\substack{i=1\dots n\\ j=1\dots p}} \in \mathbb{R}^{n\times p}$, estimated for the time period 02/01/2014 - 30/08/2019. Three steps are involved: estimation of the correlation matrix for all $p = 23$ indicators/columns of the matrix $\mathscr{X}_T$, shown in Figure 1; extraction of the factors from the correlation matrix, based on the correlation coefficients of the variables; factor rotation, in order to maximize the relationship between the variables and relevant factors.

Based on the eigenvalues criteria, three factors were selected, accounting for 91% of the total variance (see Figure 2). In order to test the sampling adequacy of the Factor Analysis, we are using the Kaiser-Meyer-Olkin (KMO) test (Kaiser (1981), Cerny and Kaiser (1977), Kaiser (1974)), which should be greater than 0.5 for a satisfactory Factor Analysis (Tabachnick and Fidell, 2013). In our sample, the KMO value is 0.92, pointing out that the Factor Analysis is suitable for structure detection. For the factor rotation,

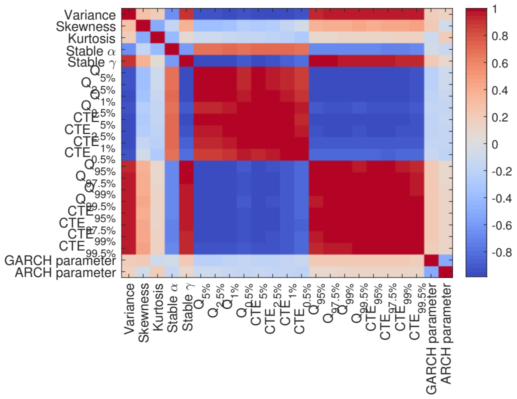

Figure 1: Correlation matrix

SFA_Cryptos

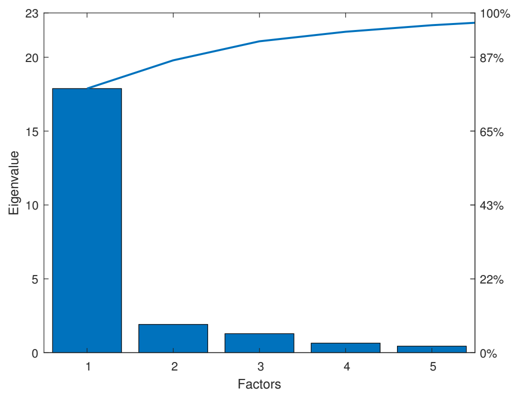

Figure 2: Scree plot

SFA_Cryptos

we used the Varimax method, which outputs orthogonal factors, while 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 Figure 3):

i. First factor: **the tail factor**, accounting for 77% of the total variance, is highly correlated with the following parameters: the tail parameter alpha and the scale parameter gamma of the α-stable distribution, the lower and upper quantiles of the distribution of log-returns, the conditional tail expectations and the variance of log-returns.

ii. Second factor: **the moment factor**, accounting for 8% of the total variance, is highly correlated with the kurtosis and the skewness coefficient of the distribution of log-returns.

iii. Third factor: **the memory factor**, accounting for 6% of the total variance, is highly correlated with the GARCH and ARCH parameters of the GARCH(1,1) model estimated for log-returns.

Based on the data revealed in Table 4, one can synthesize few characteristics of cryptocurrencies that differentiate them from the other assets. First, cryptocurrencies have higher variance of the log-return's distribution, compared to classical assets. Second, as indicated by the high values of quantiles and conditional tail expectations, cryp-

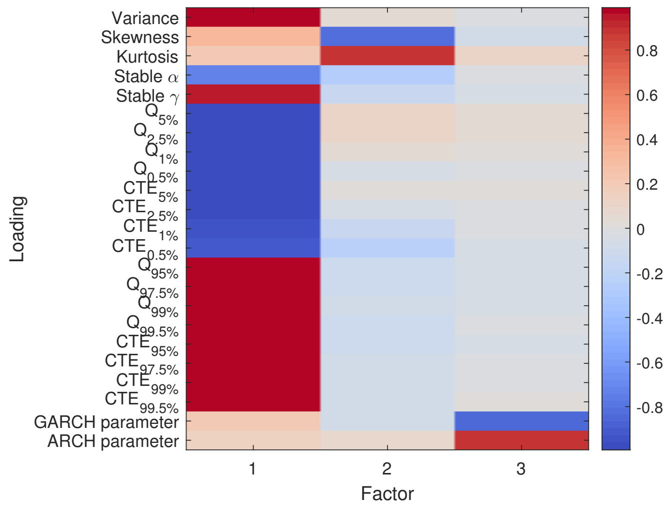

Figure 3: Loadings of the three factors

SFA_Cryptos

tocurrencies have higher propensity for extreme values, in both tails of the log-returns distribution. Third, as indicated by low values of the alpha-stable tail index, cryptocurrencies log-returns distribution has a larger departure from normality and a higher likelihood for extreme events.

Next, we map cryptocurrencies and classical assets on the 3D space defined by the factors estimated through the Factor Analysis, in order to derive some clustering effect. Figures 4 and 5 map cryptocurrencies and 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 all figures only the top 10 cryptocurrencies (according to their market capitalization) are labeled.

As shown in Figures 4 and 5, it appears to be a separation between cryptocurrencies and classical assets, mainly due to the tail factor, while the memory and moment factor are of subliminal importance.

Table 4: Assets profile based on the average values of the initial indicators, estimated for the time period 02/01/2014 - 30/08/2019

Variable	**Commodities**	**Cryptocurrencies**	**Exchange rates**	**Stocks**
$\sigma^2 \cdot 10^3$	0.365	14.563	0.028	0.270
$Skewness$	0.245	0.723	-1.233	-0.520
$Kurtosis$	22.461	28.037	38.201	13.392
$Stable_\alpha$	1.721	1.410	1.714	1.711
$Stable_\gamma$	0.010	0.047	0.003	0.009
$Q_{5\%}$	-0.027	-0.159	-0.008	-0.025
$Q_{2.5\%}$	-0.034	-0.210	-0.010	-0.033
$Q_{1\%}$	-0.044	-0.296	-0.013	-0.044
$Q_{0.5\%}$	-0.054	-0.378	-0.015	-0.054
$CTE_{5\%}$	-0.038	-0.250	-0.011	-0.038
$CTE_{2.5\%}$	-0.047	-0.319	-0.014	-0.047
$CTE_{1\%}$	-0.060	-0.428	-0.017	-0.062
$CTE_{0.5\%}$	-0.073	-0.525	-0.021	-0.076
$Q_{95\%}$	0.026	0.169	0.008	0.024
$Q_{97.5\%}$	0.034	0.243	0.010	0.030
$Q_{99\%}$	0.046	0.364	0.013	0.040
$Q_{99.5\%}$	0.057	0.480	0.015	0.049
$CTE_{95\%}$	0.039	0.297	0.011	0.034
$CTE_{97.5\%}$	0.049	0.393	0.013	0.042
$CTE_{99\%}$	0.064	0.544	0.016	0.055
$CTE_{99.5\%}$	0.080	0.671	0.018	0.066
GARCH parameter	0.706	0.796	0.728	0.637
ARCH parameter	0.118	0.159	0.078	0.130

3.3. Separating cryptocurrencies

In this section, we list the results of the methods presented in Section 2.3, in order to assess the ability of the factors produced through the Factor Analysis to separate cryptocurrencies from classical assets. First, for each of the three factors we estimated the Binary Logistic model

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

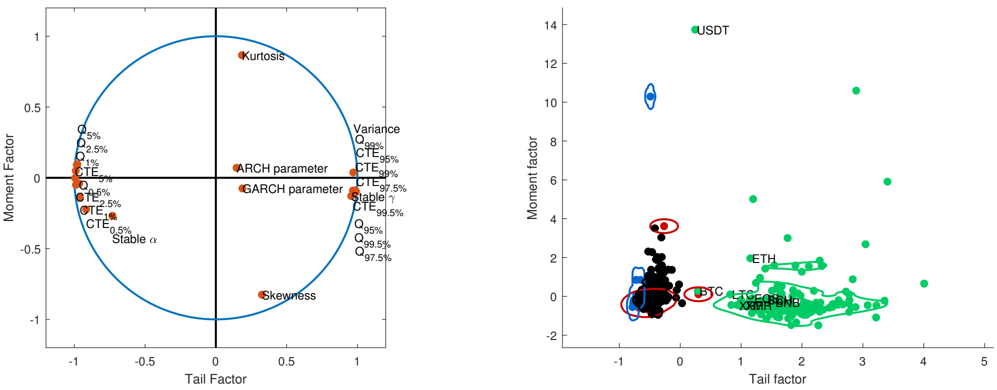

Figure 4: Loadings (left) and scores (right) based on tail and moment factor

SFA_cryptos

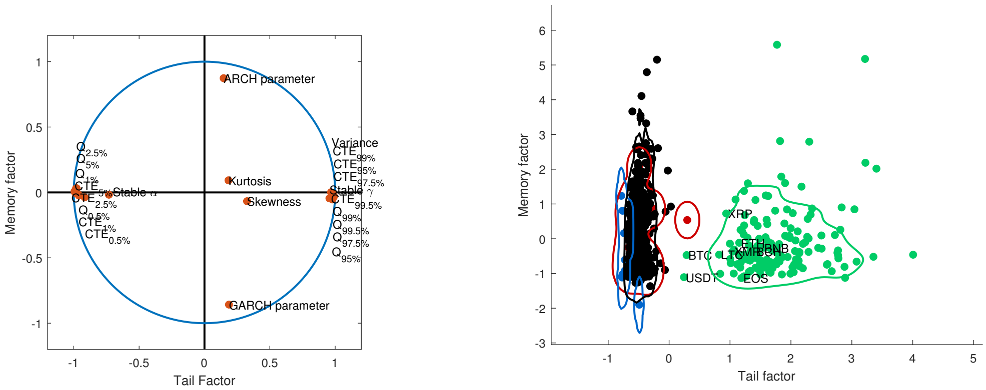

Figure 5: Loadings (left) and scores (right) based on tail and memory factor

SFA_Cryptos

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 2, with the performance measure defined by Equation A.2.

As seen in Table 5, the most important factor regarding the separation between cryptocurrencies and classical assets is the tail factor, while the other two factors have little influence. Second, we employed Discriminant Analysis and Support Vector Machines on the space defined by the two first factors (tail and moment). Figure 6 illustrates the

Table 5: Estimates of Binary logistic regression model

Exogenous factor	**Factor 1**	**Factor 2**	**Factor 3**
Estimated β_1	15.679***	-0.030	-0.084
	(3.278)	(0.077)	(0.093)
$\tilde{R}^2$	0.992	0.0003	0.002

Note: Standard errors in parentheses; *** denotes significance at 99% confidence level.

classification results using Discriminant Analysis: while Linear Discriminant Analysis fails to discriminate between cryptocurrencies and classical assets, Quadratic classifiers have a good classification power, the only cryptocurrency misclassified being Bitcoin.

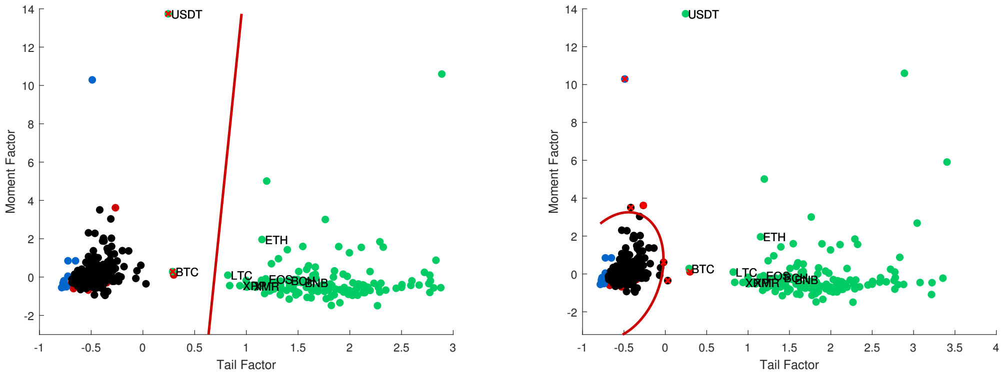

Figure 6: Discriminant Analysis: linear (left) and quadratic (right). Green dots denote cryptocurrencies, while the black dots denote the other assets; the dots highlighted in red are cases of misclassification

SFA_Cryptos

The same conclusion can be drawn by looking at the results of the Support Vector Machines non-linear classifier, according to which almost all cryptocurrencies are correctly classified using the tail factor and the moment factor (see Figure 7); the only cryptocurrency which is misclassified is Bitcoin, the overall accuracy of the classifier being 99.4%.

The k-means clustering algorithm (MacQueen, 1967) was also used, the results show-

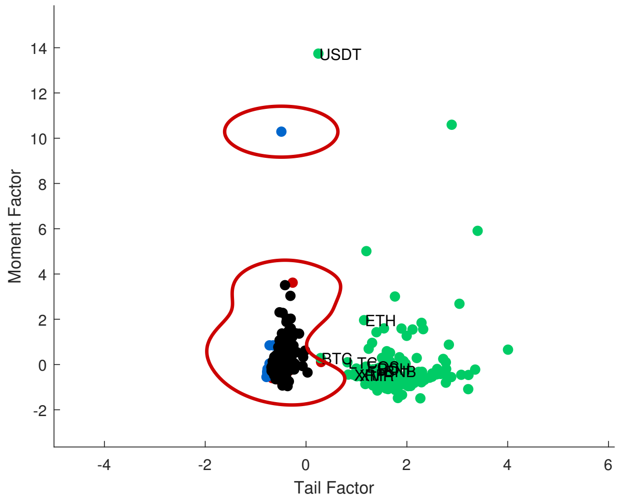

Figure 7: Support Vector Machines classification with green dots denoting cryptocurrencies, while black dots denote the other asset classes

SFA_Cryptos

ing that this method does not provide perfect classification[3] for any asset class. The optimal number of clusters, as determined by the Elbow method, is $k = 11$; however, four clusters contain only classical assets, five clusters contain only cryptocurrencies, one cluster contains Bitcoin plus classical asset (mainly stocks), while the last cluster contains Tether (USDT), Swiss Franc (CHF) and the token TenX (PAY) [4]. These clusters are shown in Figure 8, projected onto the 3D space defined by the three factors extracted through the Factor Analysis.

The projection on the 3D space defined by the Factor Analysis reveals two cryptocurrencies with atypical behaviour: Bitcoin and Tether. Thus, Bitcoin (BTC), the oldest and the most traded cryptocurrency, is closer to classical stocks and commodities, i.e. Bitcoin can be considered at the border between the classical assets and cryptocur-

[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.

[4]The complete list of the clusters components can be found at this webpage: https://github.com/QuantLet/Genus_proximum_cryptos/tree/master/Cluster_cryptos/clusters_assets.csv.

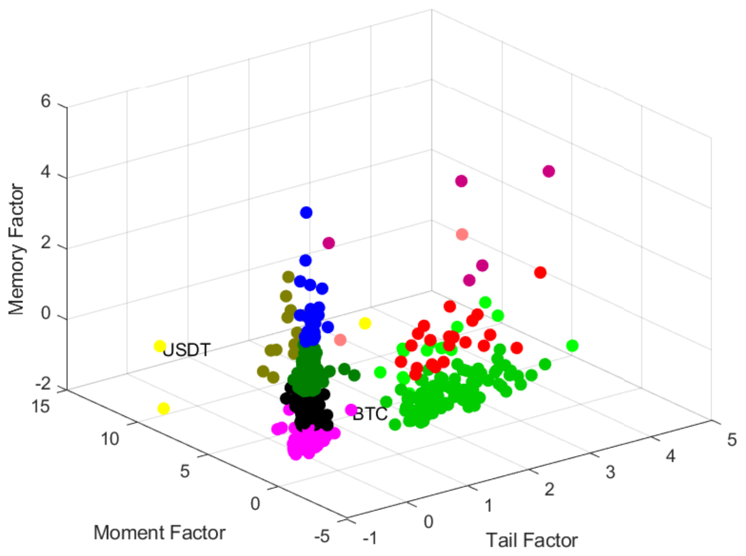

Figure 8: Projection of the clusters on the 3D space extracted trough Factor Analysis; each colour corresponds to a cluster.

Cluster_Cryptos

rencies. This result augments the findings from Dyhrberg (2016b), who concludes that Bitcoin is somewhere in between a currency (USD) and a commodity (Gold). On the other hand, Tether (USDT), a token that attempts to be tied to the US dollar, has a similar profile with the Swiss currency.

The results when applying the Maximum Variance Component Split (MVCS) method strengthen those of Binary Logistic Regression, Discriminant Analysis and Support Vector Machines, by providing complete separation of cryptocurrencies. The following notation is used: M is the number of positive equidistant angles of $[0, \pi]$ (we divide $[0, \pi]$ in M equal intervals and use the intervals left end points as projecting angles), 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 $n = 679$ is 0.014. The order of the 23 columns is as in Table 4.

In the following, we present 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.

Due to processing power constraints, we first split the data in two subsets: the first subset consists of columns 1-12 and the second includes columns 13-23. For the same reason, projection directions (A.12) are used only for $M = 3, 6$ (the number of projection directions used is M^{11} and M^{10}, respectively). Results are shown in Table 6.

Table 6: Results of the MVCS method

M	S	N_S	P_S	$minI$	$maxI$
3	1-12	43	0.024%	0.064	0.126
6	1-12	64274	0.018%	0.032	0.141
3	13-23	0	0	n/a	n/a
6	13-23	0	0	n/a	n/a

The projection direction that provided the largest index value for columns 1-12 (obtained for $M = 6$) is: $(0, 0, 0, 0.054, -0.094, 0, -0.188, 0.375, -0.25, 0.866, 0, 0)$. The index value in this direction is 0.141 and the projected values for all the assets on this projection direction are shown in Figure 9. In total 64274 (out of 362797056 tried) projection directions gave perfect classification and all provided statistically significant index values for the normal model.

The results of Table 6 indicate that columns 1-12 are more important than columns 13-23 (the largest the value of P_S for a specific value of M, the more projection directions give perfect classification, and therefore the columns used are more suitable for separating the cryptocurrencies from the other assets). Following this, we next applied the MVCS method to columns 1-12, which we further split to columns 1-6 and 7-12, for $M = 3, 6, 9, 12, 15$ and 18. For columns 7-12, no value of M provided perfect classification for the cryptocurrencies. On the other hand, for columns 1-6 and $M = 12, 15$ and 18, cryptocurrencies were completely separated from all the other assets; see Table 7.

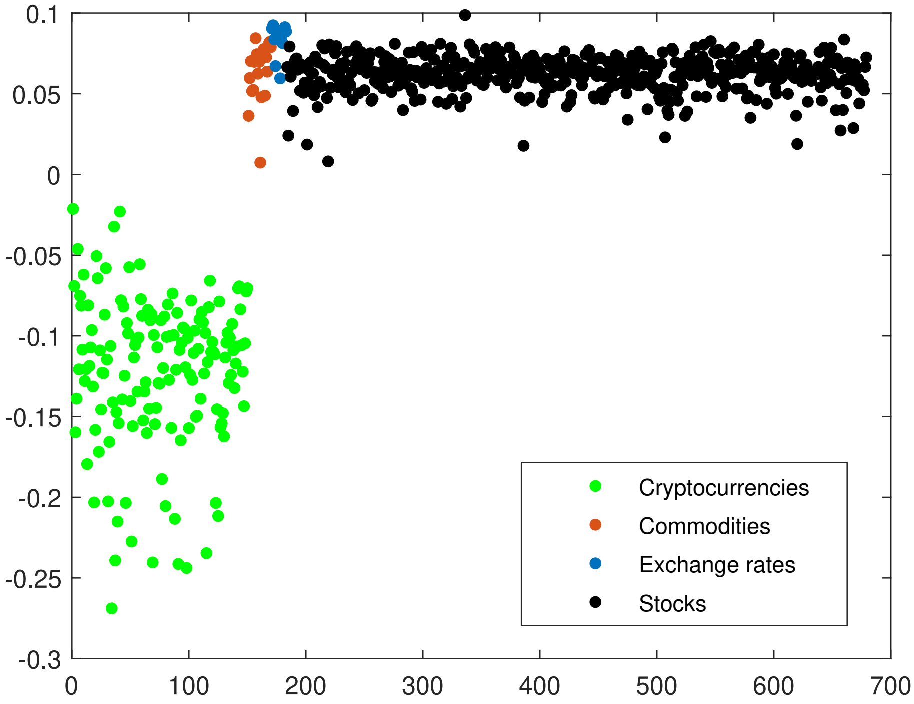

Figure 9: Projections of a subset of the data (the first 12 columns) for $M = 6$ on the projection direction that gave the largest index value among those that gave perfect classification of the cryptocurrencies.

VCS_Cryptos

Therefore, we can conclude that the most important columns for complete separation are the first six.

Next, the first six columns are further used, as they are deemed the most important, according to the above. The MVCS method is applied to all six quintets (derived by omitting in turn one of the six columns). Again, higher values of M are used (here: $M = 18, 24$ and 32), and the results being reported in Table 8. It can be concluded that the least important column is the third (Kurtosis), since its omission still provided perfect classification for the cryptocurrencies for more projection directions and for all values of M used. The second least important column is the second (Skewness). The most important columns are the first (Variance) and the fourth (α-stable parameter).

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

M	S	N_S	P_S	$minI$	$maxI$
3	1-6	0	0	n/a	n/a
6	1-6	0	0	n/a	n/a
9	1-6	0	0	n/a	n/a
12	1-6	4	0.002%	0.065	0.095
15	1-6	20	0.003%	0.066	0.101
18	1-6	2	0.0001%	0.092	0.098

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	0	0	n/a	n/a
18	1,2,3,4,6	0	0	n/a	n/a
18	1,2,3,5,6	0	0	n/a	n/a
18	1,2,4,5,6	2	0.002%	0.092	0.098
18	1,3,4,5,6	2	0.002%	0.092	0.098
18	2,3,4,5,6	0	0	n/a	n/a
24	1,2,3,4,5	0	0	n/a	n/a
24	1,2,3,4,6	0	0	n/a	n/a
24	1,2,3,5,6	0	0	n/a	n/a
24	1,2,4,5,6	37	0.011%	0.064	0.104
24	1,3,4,5,6	0	0	n/a	n/a
24	2,3,4,5,6	0	0	n/a	n/a
32	1,2,3,4,5	1	0.0001%	0.095	0.095
32	1,2,3,4,6	2	0.0002%	0.066	0.105
32	1,2,3,5,6	0	0	n/a	n/a
32	1,2,4,5,6	168	0.016%	0.060	0.112
32	1,3,4,5,6	18	0.002%	0.067	0.112
32	2,3,4,5,6	0	0	n/a	n/a

We can conclude that cryptocurrencies are financial instruments 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.

3.4. Synchronic evolution of cryptocurrencies

In order to observe the assets dynamic, we are using an expanding window approach, allowing to distinguish the evolution of the clusters. In particular, 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 matrix $\mathscr{X}_T$. 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 the expanding window approach, first, the 23-dimensional dataset is estimated for the time interval $[1, t_0]$ = [02/01/2014, 31/10/2016]; second, the time window is extended on a daily basis, up to T=30/08/2019 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[5].

Figure 10 presents a snapshot of the evolution of the assets universe using the expanding window approach[6]. Looking at the evolution of the assets universe, it appears that individual cryptocurrencies tend to develop over time similar characteristics (synchronic evolution) that make them fully distinguishable from classical assets.

In order to test this behaviour, we are using the Likelihood Ratio associated to binary logistic model 2, estimated using the expanding window approach described above. The

[5]In this approach, only the first two factors are used, as a 3D evolutionary dynamic would be difficult to read.

[6]The daily evolution of the assets universe, for the period 31/10/2016-30/08/2019, is depicted in the video *Crypto_movie*, attached to this paper as supplementary material.

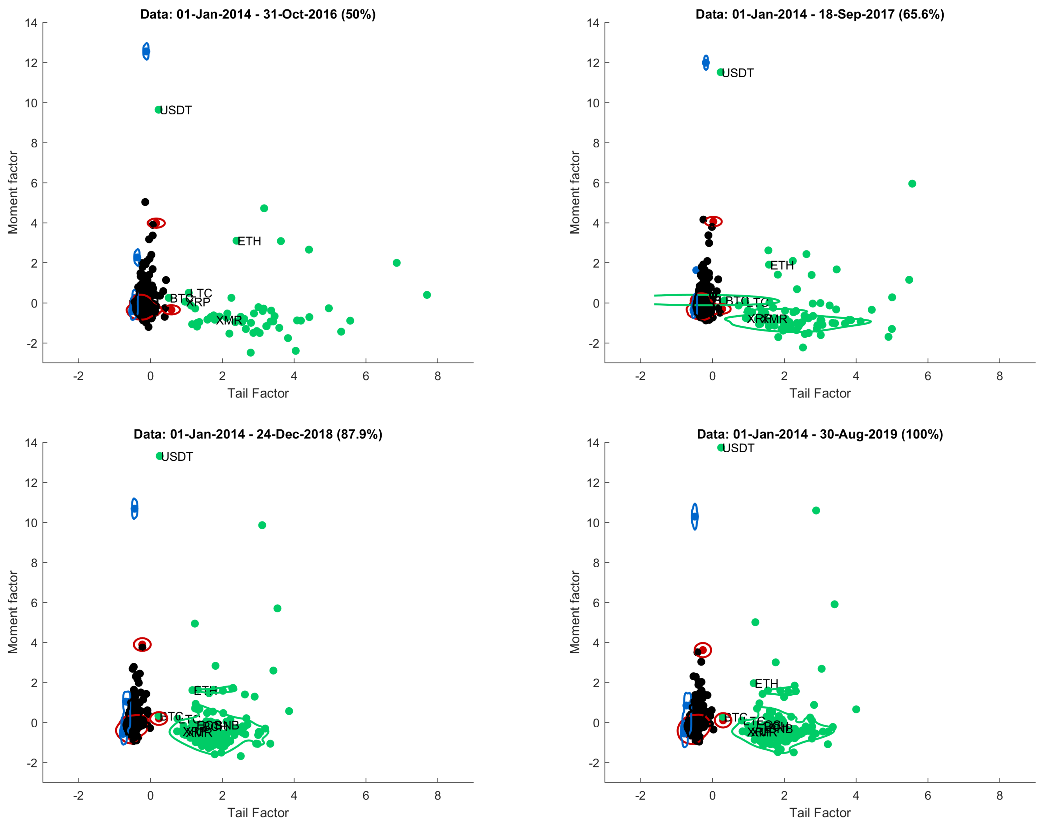

Figure 10: The evolution of the assets universe using the expanding window approach. The colour code is the following: green: cryptocurrencies, black: stocks, red: commodities, blue: exchange rates.

DFA_Cryptos

Likelihood Ratio for this model can be defined as:

$$LR(\widehat{\boldsymbol{\beta}}) = -2(\log L(\widehat{\boldsymbol{\beta}}) - \log L(\widehat{\boldsymbol{\beta}_{\boldsymbol{s}}})), \tag{3}$$

where $L(\widehat{\boldsymbol{\beta}_{\boldsymbol{s}}})$ is the maximum likelihood of a saturated model that fits perfectly the sample, while $L(\widehat{\boldsymbol{\beta}})$ is the maximum likelihood of the estimated model. In the language of Binary Logistic Regression, the Likelihood Ratio from the Equation 3 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 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

$$D = LR(\widehat{\boldsymbol{\beta}}) - LR(0), \tag{4}$$

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 $m = T - t_0 - 1 = 740$. For each model we report the Likelihood Ratio (Figure 11) and the p-value associated to Equation 4 (see Figure 12). Large p-values indicates that the model might not differ statistically from an intercept-only model.

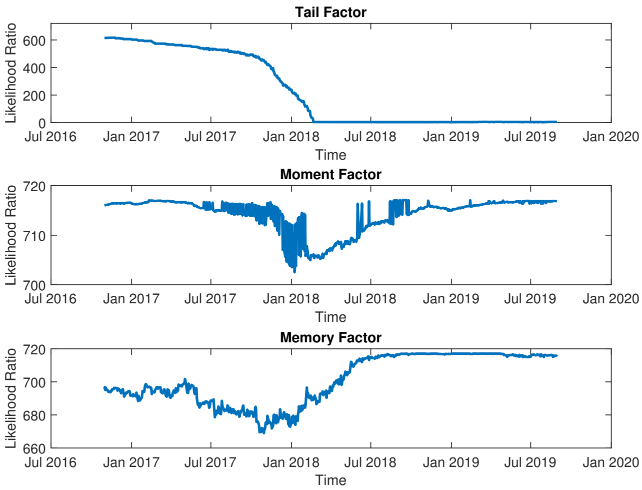

Figure 11: Likelihood Ratios for model (2), estimated on the time period 31/10/2016-30/08/2019, using an expanding window approach

CONV_Cryptos

By examining the evolution of the Likelihood Ratios, we can observe a trend change for the tail-factor-based model, starting January 2018, when the cryptocurrency market collapsed after the historical maximum of Bitcoin from December 2017. Thus, the Likelihood Ratio converges to zero, pointing out the ability of the tail factor to discriminate between cryptocurrencies and classical assets.

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

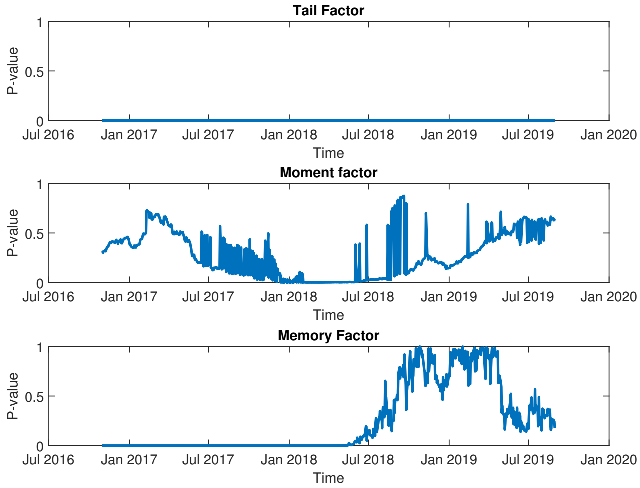

Figure 12: p-values for Equation 4, estimated on the time period 31/10/2016-30/08/2019, using an expanding window approach

CONV_Cryptos

4. Conclusions

In this paper we applied various classification techniques in order to discriminate between cryptocurrencies and classical assets, like stocks, exchange rates and commodities. Through the means of dimensionality reduction techniques and classification techniques, we proved that most of the variation among cryptocurrencies, stocks, exchange rates and commodities can be explained by three factors: the tail factor, the moment factor and the memory factor. These factors are different from the ones obtained in Liu et al. (2019) and 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.

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.

Our results help build a series of recommendations based on which researchers and practitioners can differentiate cryptocurrencies from classical assets, using one of the methods presented here. Thus, classical methods (Binary Logistic Regression, Discriminant Analysis, K-means clustering and Support Vector Machines) do not provide a complete separation of cryptocurrencies. The Maximum Variance Components Split method provides a complete separation of cryptocurrencies form the other types of assets.

By looking at the assets universe as a complex ecosystem, we provide empirical evidence that cryptocurrencies exhibit a synchronic evolution i.e. individual cryptocurrencies develop similar statistical characteristics over time, allowing them to differentiate from classical assets.

Acknowledgements

This research has received funding from the European Union's Horizon 2020 research and innovation program FIN-TECH: A Financial supervision and Technology compliance training programme under the grant agreement No. 825215 (Topic: ICT35-2018, Type of action: CSA).

Wolfgang K. Härdle thanks the supports from Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series", Czech Science Foundation(19-28231X) and the Taiwan YuShan Scholarship.

D.T. Pele would like to thank Dr. Lucian Visinescu (Texas State University), for the valuable comments and suggestions to improve the quality of the paper.

Appendix A. Classification methods

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 cryptocurrencies and classical assets. Thus, we are estimating the following family of models:

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

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 A.1, we can derive the most important factors contributing to the specific difference of cryptocurrencies. As a performance measure for Model A.1, we are using $\tilde{R}^2$ (Nagelkerke, 1991), where:

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

In Equation A.2, $L(0)$ is the maximum likelihood of the intercept-only model, $L(\widehat{\boldsymbol{\beta}})$ is the maximum likelihood of the full model, and $\widehat{\boldsymbol{\beta}}$ is the vector of Maximum Likelihood estimated parameters.

Discriminant Analysis

The aim of discriminant analysis is to classify one or more observations into *a priori* known groups, minimizing the error of misclassification (Härdle and Simar, 2019). Formally, Linear Discriminant Analysis (LDA) assumes that the input dataset is multivariate Normal: $X_i \sim \mathrm{N}(\mu_i, \Sigma)$, 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:

$$w^* = \arg\max_w \frac{|w^\top(\mu_i - \mu_j)|^2}{s_i^2 + s_j^2}, \tag{A.3}$$

$$s_i^2 = \sum_{x_i \in \omega_i} (w^\top x_i - w^\top \mu_i)^2 = w^\top S_i w, \tag{A.4}$$

giving the Linear Discriminant of Fisher (1936):

$$w^* = S_W^{-1}(\mu_i - \mu_j),\ S_W = S_i + S_j. \tag{A.5}$$

Quadratic Discriminant Analysis (QDA) follows the same procedure, but for $X_i \sim \mathrm{N}(\mu_i, \Sigma_i)$ belong to the class ω_i. In other words, 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.

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:

$$\min_{w,b} \frac{1}{2}\|w\|^2, \text{ s.t. } \quad Y_i\left(w^\top X_i + b\right) \geq 1, i = 1, \ldots, n. \tag{A.6}$$

K-means Clustering Algorithm

This clustering method was first popularized by (MacQueen, 1967), who acknowledged 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).

Variance Components Split methods: MVCS, GMVCS

These methods aim 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

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n-1}\frac{i(n-i)}{n^2}(\bar{X}_{[i+1,n]} - \bar{X}_{[1,i]})(X_{(i+1)} - X_{(i)}). \tag{A.7}$$

The summands on the right side of Equation A.7 measure between-groups variations. The standardized sample variance components

$$W_i = W_i(X_1, \ldots, X_n) \tag{A.8}$$

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

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

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

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

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

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, 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

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

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

The number of projection directions to be used is M^{r-1}. The method is thus 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 A.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 $\max\{\mathcal{I}_{\mathcal{X}}(\mathbf{a}), \mathbf{a} \in \mathcal{D}\}$, determines two clusters. In practice, its approximation is obtained using $\mathcal{D}_M$. The splitting of these clusters may continue (Yatracos, 2013).

References

C. Alexander and M. Dakos. A critical investigation of cryptocurrency data and analysis. *Quantitative Finance*, 20(2):173–188, 2020. ISSN 1469-7688. doi: 10.1080/14697688.2019.1641347.

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.

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_3.

C. Conrad, A. Custovic, and E. Ghysels. Long- and Short-Term Cryptocurrency Volatility Components: A GARCH-MIDAS Analysis. *Journal of Risk and Financial Management*, 11(2):23, 2018. doi: 10.3390/jrfm11020023.

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.

F. A. Enoksen, C. Landsnes, K. Lučivjanská, and P. Molnár. Understanding risk of bubbles in cryptocurrencies. *Journal of Economic Behavior & Organization*, 176: 129–144, 2020. ISSN 01672681. doi: 10.1016/j.jebo.2020.05.005.

E. F. Fama and K. R. French. Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance*, 51(1):55–84, 1996. ISSN 00221082. doi: 10.1111/j.1540-6261.1996.tb05202.x.

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.

T. Fischer, C. Krauss, and A. Deinert. Statistical Arbitrage in Cryptocurrency Markets. *Journal of Risk and Financial Management*, 12(1):31, 2019. doi: 10.3390/jrfm12010031.

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, 5. ed. edition, 2019. ISBN 9783030260057.

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.

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

I.M.F. Treatment of Crypto Assets in Macroeconomic Statistics. *International Monetary Fund, Statistics Department*, 2019. URL https://www.imf.org/external/pubs/ft/bop/2019/pdf/Clarification0422.pdf.

H. Jiang and L. Han. Does Income Diversification Benefit the Sustainable Development of Chinese Listed Banks? Analysis Based on Entropy and the Herfindahl-Hirschman Index. *ENTROPY*, 20(4), 2018. ISSN 1099-4300. doi: 10.3390/e20040255.

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.-Y. Li, Q. Cai, G.-F. Gu, and W.-X. Zhou. Exponentially decayed double power-law distribution of Bitcoin trade sizes. *PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS*, 535:122380, 2019. ISSN 0378-4371. doi: 10.1016/j.physa.2019.122380.

Y. Liu and A. Tsyvinski. Risks and Returns of Cryptocurrency. *NBER Working Paper No. 24877*, 2018. doi: 10.3386/w24877.

Y. Liu, A. Tsyvinski, and X. Wu. Common Risk Factors in Cryptocurrency. *NBER Working Paper No. 25882*, 2019. doi: 10.3386/w25882.

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.

S. A. Manavi, G. Jafari, S. Rouhani, and M. Ausloos. Demythifying the belief in cryptocurrencies decentralized aspects. A study of cryptocurrencies time cross-correlations with common currencies, commodities and financial indices. *PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS*, 556:124759, 2020. ISSN 0378-4371. doi: 10.1016/j.physa.2020.124759.

M. Mirtaheri, S. Abu-El-Haija, F. Morstatter, G. V. Steeg, and A. Galstyan. Identifying and Analyzing Cryptocurrency Manipulations in Social Media. 2009. URL http://arxiv.org/pdf/1902.03110v1.

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.

J. A. Núñez, M. I. Contreras-Valdez, and C. A. Franco-Ruiz. Statistical analysis of Bitcoin during explosive behavior periods. *PloS one*, 14(3):e0213919, 2019. doi: 10.1371/journal.pone.0213919.

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

M. Schnaubelt, J. Rende, and C. Krauss. Testing Stylized Facts of Bitcoin Limit Order Books. *Journal of Risk and Financial Management*, 12(1):25, 2019. doi: 10.3390/jrfm12010025.

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.

C. Trucíos, A. K. Tiwari, and F. Alqahtani. Value-at-risk and expected shortfall in cryptocurrencies' portfolio: a vine copula–based approach. *APPLIED ECONOMICS*, 52(24):2580–2593, 2020. ISSN 0003-6846. doi: 10.1080/00036846.2019.1693023.

M. Veillete. STBL: Alpha stable distributions for MATLAB. *Matlab Central File Exchange*, 2012. URL https://www.mathworks.com/matlabcentral/fileexchange/37514-stbl-alpha-stable-distributions-for-matlab.

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.