This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 825215. All material presented here reflects only the authors' view. The European Commission is not responsible for any use that may be made of the information it contains.

Blockchain Use Cases

FIN-TECH HO2020

February 7, 2020

Table of contents

Use Case I: Libra or Librae? Basket-based Stablecoins (Paolo Giudici, Thomas Leach and Paolo Pagnottoni, UNIPV)

Background

- The term stablecoin is a by-product of the cryptoasset domain. Cryptoasset's inherently inefficient design constrain their ability to serve as digital money per se and has ultimately limited their rate of adoption ((see Chiu and Koeppl (2017); Schilling and Uhlig (2019))
- The first documented attempt to deal with the matter of price volatility in cryptoassets is BitShares, refereed by [**?**] as "a new type of financial product called a Polymorphic Digital Asset [PDA] that can track the value of gold, silver, dollars, or other currencies."
- Systemic implications of cryptoassets have largely been classed as manageable or limited as linkages with financial markets and the real economy remain relatively low: Manaa et al. (2019), Giudici and Abu-Hashish (2018), Giudici and Pagnottoni (2019).
- Facebook's Libra has pushed stablecoins up the agenda for regulators and supervisors. Facebook can push Libra to its vast user-base, approximately 2.41 billion monthly active users.[1]

[1] https://newsroom.fb.com/company-info/

Taxonomy of stablecoins

Based on Bullmann, 2019

- **Tokenised funds** - denote stablecoins that are a claim on a pool of collateral that consists of funds, including cash, electronic money, commercial bank money or central bank reserve deposits e.g. Tether, Utility Settlement Coin
- **Off-Ledger Collateralised** - stablecoins that are a claim on a pool of collateral that is comprised of various assets e.g. multiple currencies, T-Bills etc
- **On-Ledger Collateralised** - stablecoins that are a claim on a pool of underlying collateral that is held on a blockchain e.g. Dai
- **Algorithmic** - take users expectations into account to stabilise the value of the coin (mostly conceptual) e.g. BasisCoin

Aim of the paper

- Can a basket based stable coin function as a global e-currency ?
- What is the optimal way to construct a stablecoin whose value is derived from a basket of currencies?
- Is a basket based stable coin ("Librae") better than single currency based stablecoins ("Libra"), particularly from the viewpoint of remittances? in 2018 overall global remittance grew 10% to 689 billion dollars, including 529 billion dollars to low income countries. India the largest, followed by China, the Philippines, Mexico and Nigeria.
- Which currencies lead volatility spillovers on the others?
- How do shocks to the values of the leading currencies affect the stability of basket-based and single currency based stablecoins?

Reduced Normalized Values

- Hovanov et al. (2004) show that the values of any given currency depend on the base currency chosen
- $\Rightarrow$ this creates ambiguity in the valuation of a currency and makes it difficult to examine the dynamics of the time series of currency values
- To overcome this base currency problem they proposed a reduced (to the moment t_0) normalized value in exchange of the i-th currency:

$$\mathrm{RNVAL}_i\left(t/t_0\right) = \frac{c_{ij}(t)}{\sqrt[n]{\prod_{k=1}^{n} c_{kj}(t)}} \Big/ \frac{c_{ij}\left(t_0\right)}{\sqrt[n]{\prod_{k=1}^{n} c_{kj}\left(t_0\right)}} = \sqrt[n]{\prod_{k=1}^{n} \frac{c_{ik}(t)}{c_{ik}\left(t_0\right)}}$$

Optimal basket weights

- The RNVAL allows the computation of a unique optimal, minimum variance currency basket regardless of the base currency choice
- The derivation of the minimum variance currency basket is calculated by searching the optimal weight vector w^* that solves the following optimal control problem:

$$\text{Min}\left(S^2(w) = \sum_{i,j=1}^{n} w_i w_j \operatorname{cov}(i,j) = \sum_{i=1}^{n} w_i^2 s_i^2 + 2\sum_{i,j=1}^{n} w_i w_j \operatorname{cov}(i,j) \right)$$

subject to

$$\begin{cases} \sum_{i=1}^{n} w_i = 1 \\ \qquad\quad w_i \geq 0 \end{cases}$$

Impulse Response Functions and Spillovers

- To determine the impact of shocks on the stablecoins we start from estimating a Vector Autoregressive model, i.e. :

$$x_t = \sum_{i=1}^{k} \Phi_i x_{t-i} + \varepsilon_t$$

  - Φ_i :$(n \times n)$ VAR parameter matrices
  - k : autoregressive order
  - ε_t : zero-mean white noise process having variance-covariance matrix Σ_ε
- We take the differences of the reduced normalised values (stationarity)
- We then analyse impulse response functions (IRFs) and spillovers (Diebold and Yilmaz, 2012) in order to retrieve how a unit shock in one currency impacts the stablecoins

Data

- FX pairs for the period Jan 2002 - Nov 2019 (daily observations)
  - USD, EUR, JPY, CNY, GBP
  - INR (india) , MXN (Mexico), NGN (Nigeria), PHP (Philippines)

Optimal weights

Currency	USD	CNY	EUR	GBP	JPY
Optimal Weights	0.21	0.14	0.21	0.21	0.23
IMF SDR Weights	0.42	0.11	0.31	0.08	0.08

Table 1: Weights of the currency in the chosen basket, according to our methodology (Optimal) and the IMF Special Drawing Rights (IMF SDR)

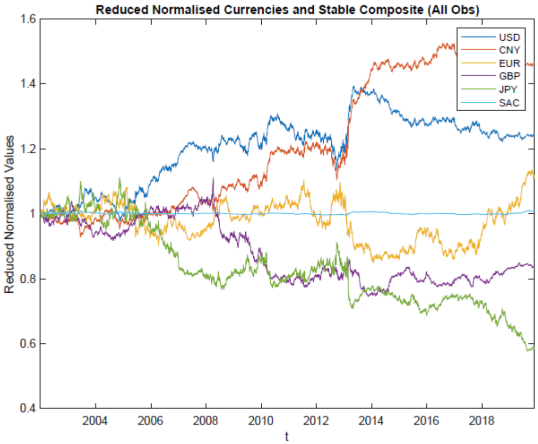


Figure 1: Time evolution of the Reduced Normalised VALue of the basket currencies (USD, CNY, EUR,GBP, JPY), and of the basket based stable coin (SAC)

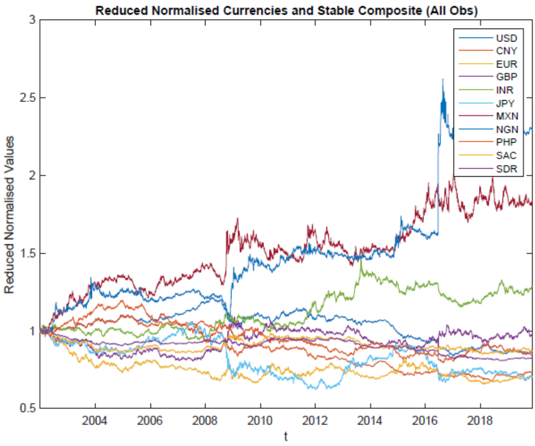


Figure 2: Time evolution of the Reduced Normalised VALue of the basket currencies (USD, CNY, EUR,GBP, JPY), of the considered emerging market currencies (INR, MXN, NGN, PHP) and of the basket based stable coins (SAC, SDR)

	USD	CNY	EUR	GBP	JPY	SAC
USD	1	0.79	-0.48	-0.76	-0.86	0.023
CNY	0.78	1	-0.45	-0.83	-0.86	0.012
EUR	-0.48	-0.45	1	0.2	0.24	0.04
GBP	-0.76	-0.83	0.22	1	0.66	0.027
JPY	-0.86	-0.86	0.24	0.66	1	0.02
SAC	0.02	0.01	0.039	0.027	0.02	1
σ	0.11	0.2	0.06	0.09	0.12	0.003

Table 2: Volatility and Correlations between the RNVALs of the basket currencies, and the optimal basket based stable coin.

	USD	CNY	EUR	GBP	INR	JPY	MXN	NGN	PHP	SAC	SDR
σ_{all}	0.09	0.14	0.07	0.06	0.13	0.11	0.22	0.41	0.10	0.04	0.05
σ_{pre}	0.05	0.02	0.04	0.03	0.04	0.07	0.11	0.35	0.03	0.01	0.03
σ_{cri}	0.02	0.04	0.03	0.03	0.12	0.07	0.05	0.04	0.03	0.02	0.02
σ_{post}	0.05	0.06	0.08	0.07	0.04	0.08	0.15	0.09	0.07	0.03	0.02

Table 3: Volatility of the RNVALs of the basket currencies, of the emerging market currencies, and of the two basket based stable coins, over the whole period (all), the pre-crisis period (pre), the crisis period (cri) and the post-crisis period (post).

	USD	CNY	EUR	GBP	JPY	FROM
USD	44.94	35.33	13.02	6.67	0.04	11.01
CNY	34.49	49.40	10.76	5.34	0.00	10.12
EUR	15.81	15.22	62.29	6.48	0.19	7.54
GBP	11.4	10.21	6.28	69.58	2.53	6.08
JPY	0.41	0.14	0.01	3.94	95.51	0.90
TO	12.42	12.18	6.01	4.49	0.55	35.66

Table 4: Spillover table

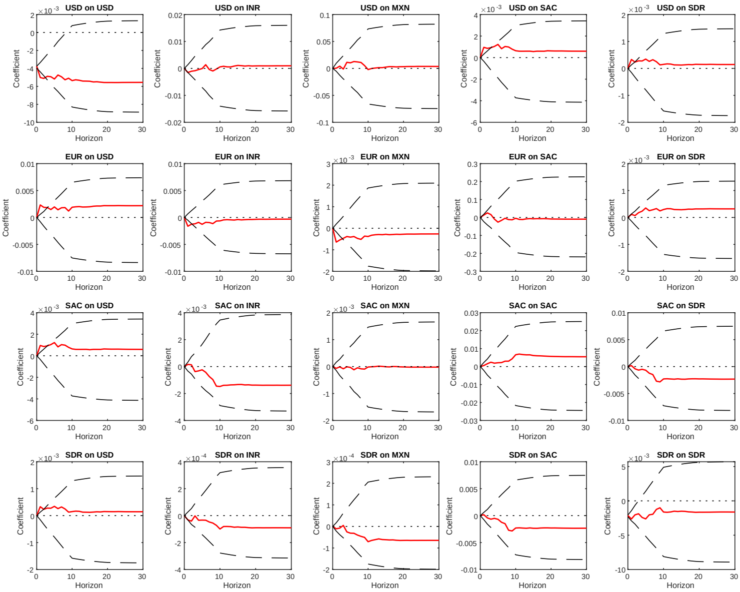


Figure 3: Cumulative impulse response functions

Conclusions

- We have shown how to construct a basket based stablecoin based on an optimal control problem to determine the weights
- The basket based stable coin is less volatile than single currencies, and it may be attractive particularly for remittances.
- The basket based stable coin is more resilient under single currency shocks

Use Case II: ICOs success drivers: a textual and statistical analysis (Paola Cerchiello and Anca Mirela Toma, UNIPV)

Fraud detection - ICO case study

- Initial Coin Offerings are a new yet uncovered mean to raise funds through tokens at the interplay of **crowdfunding** and **blockchain**.
- The acronym stands for initial offering of any crypto asset (ESMA).
- Few numbers (based on Coinschedule.com)
  - around **6** bi USD raised in 2017 by **456** ICOs
  - around **21.7** bi USD raised till the end of 2018 by **1076** ICOs
  - around **2.4** bi USD raised till April 2019 by **60** ICOs
- The crypto assets bring benefits but also risks due to the presence of criminal activity.
- Financial market authorities are very prudent and some countries ban straightaway all ICOs from their jurisdiction.

ICO risks and regulation activities

- In the current hype, the risks of ICOs are a dangerous cocktail
  - Exaggeration of expected returns
  - The knowledge and expertise required is underestimated
  - Lack of transparency
  - Market driven by speculation and even manipulation
- The crypto assets bring benefits but also risks due to the presence of criminal activity.
- Financial market authorities are very prudent and some countries ban straightaway all ICOs from their jurisdiction.

Do we need to regulate ICOs?

- Regulation activities started in 2017 with different typologies of advises from the institutions, covering the potential and the mechanisms of the finance mechanism.
- Worldwide jurisdictions have opted for one of the following three solutions for ICO regulation:
  - proactive approach
  - careful consideration
  - undefined approach
  - Dura lex sed lex

ICOs peculiarities: Success, Failure or Scam

- The success of ICOs relays on the decentralized nature of P2P technology and on the lack of regulation.
- By February 2018 almost half of ICOs sold in 2017 failed (Hankin, 2018).
- Recent scientific studies are few, relying mainly on financial data or on the legal side (Adhami et al, Zeztche et al.)
- Our main goal is to contribute with an ensemble of alternative data and statistical approaches to the jigsaw puzzle of alternative crowdfunding systems, detecting which characteristics of an **ICO are significantly related to success and fraudulent behaviours.**

Data

- Data collecting process involves structured and unstructured information
  - websites specialized in providing financial information and in listing the existing ICOs (icobench.com, TokenData.io, ICODrops.com, CoinDesk.com)
  - Telegram social channel
- Typology of data collected:
  - categorical, numerical and textual data.
  - characteristics of white papers: elicited through textual analysis;
  - team members: quantitative and qualitative information;
  - type of business;
  - geographical distribution;
  - the supporting community: social channels;
  - Telegram's chat text.

Methodology - Response Variable

The response variable representing the status of an ICO is made up of 3 classes, intended as follows:

- **Success**: the given ICO collects the predefined cap within the time horizon of the campaign;
- **Failure**: the given ICO does not collect the predefined cap within the time horizon of the campaign;
- **Scam**: the given ICO is discovered to be a fraudulent activity during the campaign and described as such by all the platforms we use for data gathering (namely ICObench and Telegram).

Methodology - Explanatory variables

Table 5: Employed Covariates

class0	f=failed, sc=scam su=success
class1	0=success, 1=scam
class2	0=failed, 1= success
w_site	Website (dummy)
tm	Telegram (dummy)
w_paper	White paper (dummy)
usd	presale price in USD
tw	Twitter (dummy)
fb	Facebook (dummy)
ln	Linkedin (dummy)
yt	Youtube (dummy)
gith	Github (dummy)
slack	Slack (dummy)
reddit	Reddit (dummy)
btalk	Bitcointalk (dummy)
mm	Medium (dummy)
nr_team	Number of Team members
adv	Existence of advisors (dummy)
nr_adv	Number of advisors
project	Official name of the ICO
nr_tm	Number of users in Telegram
tot_token	Number of Total Tokens
Pos_Bing	Standardized number of positive words for BL list
Neg_Bing	Standardized number of negative words for BL list
Sent_Bing	Standardized sentiment for BL list
Pos_NRC	Standardized number of positive words for NRC list
Neg_NRC	Standardized number of negative words for NRC list
Sent_NRC	Standardized sentiment for NRC list

Methodology

- Using supervised classification models we will get insights for discriminating and classifying ICOs by their probability of success.
- At the same time, text mining methods will be the tools for dealing with the large corpus of text coming from the Telegram chats and the white papers.

Analysis – I Logistic Regression for Successful ICOs

Logistic regression aims at classifying the dependent variable into two groups, characterized by a different status [1=success vs 0=scam or 1=success vs 0=failure]:

$$ln(\frac{p_i}{1-p_i}) = \alpha + \sum_j \beta_j x_{ij}, \tag{1}$$

where p_i is the probability of the event of interest, for ICO i, $x_i = (x_{i1}, \ldots, x_{ij}, \ldots, x_{iJ})$ the J covariates from which the probability of success (or scam) can be obtained as:

$$p_i = \frac{1}{1 + exp(\alpha + \sum_j \beta_j x_{ij})}, \tag{2}$$

Analysis – II Multilogit Regression

Since the target variable is naturally categorized according to 3 classes, success, failure and scam we extend the aforementioned binary logistic regression to a multinomial one. Such model assesses all the categories of interest at the same time as follows:

$$ln(\frac{p_k}{1 - p_K}) = \alpha_k + \sum_j \beta_k x_{ij}, \tag{3}$$

where p_k is the probability of kth class for $k = 1, ..., K$ given the constraint that $\sum_K p_k = 1$.

Textual Analysis

We have applied a Bag of Word (BoW) approach, where a text is represented as an unordered collection of words, considering only their counts in each comment of the chat.

The word and document vectorization has been carried out by creating a Term Document Matrix (TDM).

Classical text cleaning procedures have been put in place like: stop-words, punctuation, unnecessary symbols and space removal, specific topic words addition.

For descriptive purposes we have used wordclouds for each and every Telegram chat according to the general content and to specific subcategories like sentiments and expressed moods.

Textual analysis – I

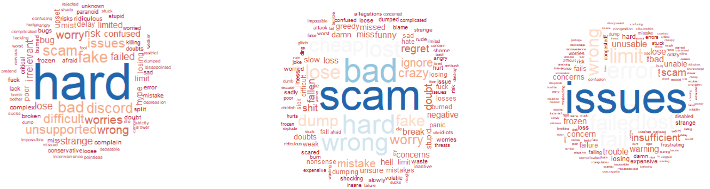


Figure 4: Wordcloud with negative words – failed – scam – success

Textual analysis – II

Figure 5: Wordcloud with positive words – failed – scam – success

Sentiment Analysis

We decided to focus on a dictionary based approach, adapting appropriate lists of positive and negative words relevant to ICOs topics in English language. We employ 3 vocabularies from the R package 'tidytext':

- AFINN from Finn Arup Nielsen;
- BING from Bing Liu and collaborators;
- NRC from Saif Mohammad and Peter Turney.

By applying the above lexicons, we produce for each and every ICO a sentiment score as well as counts for positive and negative words. All these indexes are used as additional predictors within the logistic models.

Sentiment Analysis - II

Figure 6: NRC emotion lexicon – success – scam

Results - I

Table 6: Results from Logistic regression on Success/Failure

	Dependent variable:
	class2
tw	2.481^{*}
	(1.381)
Paper_du	1.351^{**}
	(0.635)
nr_adv	0.461^{***}
	(0.135)
nr_team	0.233^{***}
	(0.088)
Sent_NRC_sc	2.187^{***}
	(0.595)
Constant	-3.601^{**}
	(1.458)
Observations	196
Akaike Inf. Crit.	89.41
McFadden pseudo R^2	0.63
McFadden Adj. pseudo R^2	0.57
Cox & Snell pseudo R^2	0.49
Note:	*p<0.1; **p<0.05; ***p<0.01

Results – II

Table 7: Results from multilogit regression: failure and scam compared to success

	Dependent variable:	
	f	sc
	(1)	(2)
Oweb_dum	-1.962^{**}	0.093
	(0.977)	(0.773)
adv_dum	-0.899	-1.707^{***}
	(0.809)	(0.571)
Paper_du	-0.728	-2.158^{***}
	(0.915)	(0.657)
Sent_NRC_sc	-1.390^{*}	-2.606^{***}
	(0.731)	(0.703)
Constant	-0.628	-0.572
	(0.997)	(0.925)
Akaike Inf. Crit.	166.339	166.339
Pseudo R square	McFadden 0.43 - McFadden Adj. 0.36- Cox & Snell 0.44	
Note:		*p<0.1; **p<0.05; ***p<0.01

Variance inflation factor

Table 8: VIF index for logistic regression model

tw	Paper_du	nr_adv	nr_team	Sent_NRC_sc
1.229	1.033	1.067	1.053	1.228

Table 9: VIF index for multilogit regression model

Oweb_dum	adv_dum	Paper_du	Sent_NRC_sc
6.395	2.207	3.822	7.034

Preliminary conclusions and ongoing research

From the logistic regression the relevant variables are: the presence of a white paper, of a Twitter account, number of elements of the team, number of advisors, and scaled sentiment score.

From text analysis: the net sentiment based on NRC lexicon has a positive impact in discriminating success ICOs from failure and scam ones.

From the multilogit regression we report results for fraudulent and scam ICOs compared to successful ones.

Preliminary conclusions and ongoing research

This paper represents a preliminary work and we are running a more detailed and complete NLP analysis by:

- increasing the size of the sample bu using the API access to the IcoBench Platform, and therefore analyzing the 5000 projects published there.
- refining the sentiment analysis and the dictionary based method.
- through topic modelling we aim at producing a quality index for white-paper to be included in the classification models, as a possible driver of success and/or scam activity.

Use Case III: Phenotypic convergence of Cryptocurrencies (Niels Wesselhöfft, Wolfgang K. Härdle, Yannis Yatracos, Daniel Traian Pele, Michalis Kolossiatis; Humboldt Universität zu Berlin, Bucharest University of Economic Studies and University of Cyprus)

Genus differentia approach

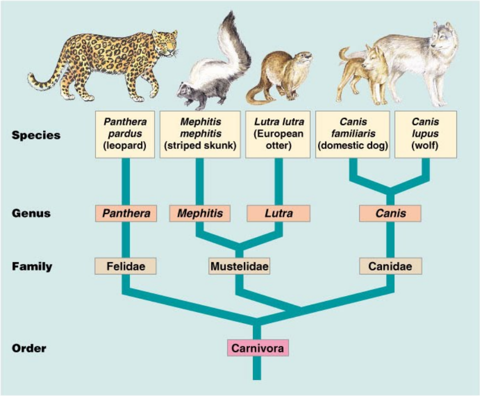


Figure 7: Genus differentia approach in biology

Genus differentia approach

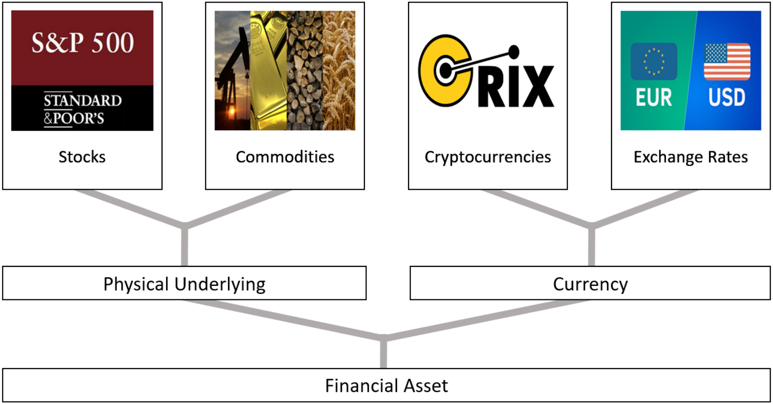


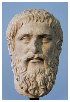
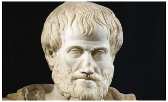
Figure 8: Genus differentia approach in finance

Aim of classification

- Genotypic differentiation
  - Biology - the change in DNA sequences.
  - Finance - the underlying process of price manifestation.
- Phenotypic differentiation
  - Biology - classification based on behavior and features of a species.
  - Finance - classification based on statistical features of the price series.

Motivation

- Question: What defines cryptocurrencies?

- Plato: man is an upright, featherless biped, with broad, fat nails.
- Aristotle: definition of a species consists of genus proximum and differentia specifica.
- Goal: Define cryptocurrencies in terms of their genus proximum and differentia specifica.
- Method: Find latent variables, to form groups of shared characteristics.
- Finding: Phenotypic convergence of cryptocurrencies, i.e. asymptotic speciation.
- Implication: Cryptocurrencies are a different species in the ecosystem of financial instruments.

Literature review

- Dyhrberg (2016): BTC has similarities to both GOLD and the USD, being in between a currency and a commodity.
- Baur et al. (2018): BTC volatility and correlation characteristics are distinctively different compared to GOLD and USD.
- Härdle et al. (2018): BTC, XRP, LTC, ETH returns exhibit higher volatility, skewness and kurtosis compared to GOLD and S&P500 daily returns.
- Henriques et al. (2018): BTC can serve as a substitute for GOLD in a portfolio.
- Zhang et al. (2018): Cryptocurrencies presents heavier tails and higher Hurst exponent than the classical assets.

Data

- Sample: $n = 544$ assets.
- New asset class
  - Cryptocurrencies (CRIX): $n_1 = 14$ List
- Old asset classes
  - Stocks (S&P 500): $n_2 = 497$
  - Exchange rates: $n_3 = 13$ List
  - Commodities (Bloomberg Commodity Index): $n_4 = 20$ List
- Daily data from 2014-10-22 to 2018-10-16 (4 years of daily trading data).

CRIX components

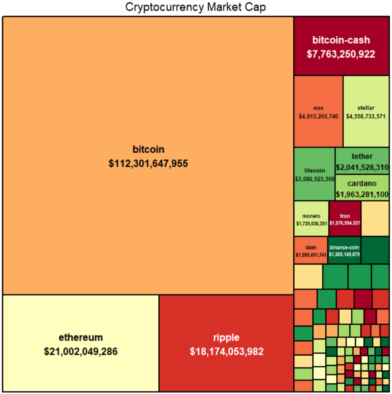


Figure 9: Components of the CRIX cryptocurrency index. Mkt_cryptos

Statistical assessment

- Return X is a r.v. with cdf $F()$ from which $p = 23$ statistics are estimated.
- Moments of order $k \in \mathbb{R}^+$, $\mu_k = \mathrm{E}\{(X-\mu)^k\}$.
  - variance: $\sigma^2 = E\left\{(X-\mu)^2\right\}$;
  - skewness: $Skewness = E\left\{(X-\mu)^3\right\}/\sigma^3$;
  - kurtosis: $Kurtosis = E\left\{(X-\mu)^4\right\}/\sigma^4$.
- Tails: $\alpha \in \{0.005, 0.01, 0.025, 0.05, 0.95, 0.975, 0.99, 0.995\}$.
  - $Q_\alpha = \inf\{x \in \mathbb{R} : \alpha \leq F(x)\}$;
  - $CTE_\alpha = \begin{cases} \mathrm{E}\{X \mid X < Q_\alpha\}, & \alpha < 0.5 \\ \mathrm{E}\{X \mid X > Q_\alpha\}, & \alpha > 0.5 \end{cases}$
- Scaling and memory parameters
  - Alpha-stability Alpha-stability
  - Autocorrelation (Pearson correlation)
  - Long memory (Hurst parameter)

Assets profile

Factor	Estimate	Cryptos	Stocks	Commodities	Exchange Rate	Bitcoin
Tail factor	$\sigma^2 \cdot 10^3$	7.88	0.28	0.37	0.03	1.50
	S_α	1.44	1.70	1.75	1.76	1.32
	$S_\gamma \cdot 10^3$	36.76	8.73	9.85	3.17	16.02
	$Q_{0.5\%}$	−0.26	−0.06	−0.05	−0.02	−0.14
	$Q_{1\%}$	−0.22	−0.04	−0.04	−0.01	−0.11
	$Q_{2.5\%}$	−0.15	−0.03	−0.03	−0.01	−0.09
	$Q_{5\%}$	−0.11	−0.02	−0.03	−0.01	−0.06
	$Q_{95\%}$	0.13	0.02	0.03	0.01	0.06
	$Q_{97.5\%}$	0.20	0.03	0.04	0.01	0.08
	$Q_{99\%}$	0.29	0.04	0.05	0.01	0.11
	$Q_{99.5\%}$	0.38	0.05	0.06	0.02	0.14
	$CTE_{0.5\%}$	−0.33	−0.08	−0.07	−0.02	−0.18
	$CTE_{1\%}$	−0.28	−0.06	−0.06	−0.02	−0.15
	$CTE_{2.5\%}$	−0.22	−0.05	−0.05	−0.01	−0.12
	$CTE_{5\%}$	−0.17	−0.04	−0.04	−0.01	−0.10
	$CTE_{95\%}$	0.23	0.04	0.04	0.01	0.09
	$CTE_{97.5\%}$	0.31	0.04	0.05	0.01	0.12
	$CTE_{99\%}$	0.41	0.06	0.07	0.02	0.15
	$CTE_{99.5\%}$	0.50	0.07	0.08	0.02	0.18
Moment factor	*Skewness*	0.97	−0.51	0.29	−1.22	−0.28
	Kurtosis	20.35	12.92	20.72	33.99	8.58
Memory factor	$\rho(1) \cdot 10^3$	40.63	−2.16	−13.18	−11.45	16.64
	H	0.57	0.51	0.53	0.51	0.57

Table 10: Assets profile

Factor analysis

- Estimate the correlation matrix for all variables.
- Factor extraction based on the correlation of the coefficients.
- Factor rotation.

Correlation matrix

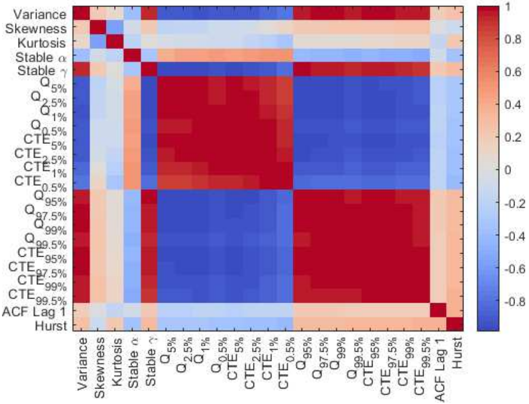


Figure 10: Correlation matrix of the statistical estimates. SFA_cryptos

Factor model

- Linear Factor model

$$X = QF + \mu + \varepsilon,\ \varepsilon \sim G() \tag{4}$$

  - X is the initial matrix of p variables
  - Q is a matrix of the non-random loadings
  - F are the common k factors ($k < p$)
  - μ is the vector of the means of initial p variables
  - ε is a matrix of the random specific factors
  - Random vectors F and U are unobservable and uncorrelated

Factor model extensions

- Time-varying factor model

$$X_t = Q_t F_t + \mu_t + \varepsilon_t,\ \varepsilon_t \sim G() \tag{5}$$

- Nonlinearities in the factors

$$X = Qm(F) + \mu + \varepsilon,\ \varepsilon \sim G() \tag{6}$$

- General nonlinear

$$X = m(F) + \varepsilon,\ \varepsilon \sim G(), \tag{7}$$

where $m()$ is a function

Factors loadings and scree plot

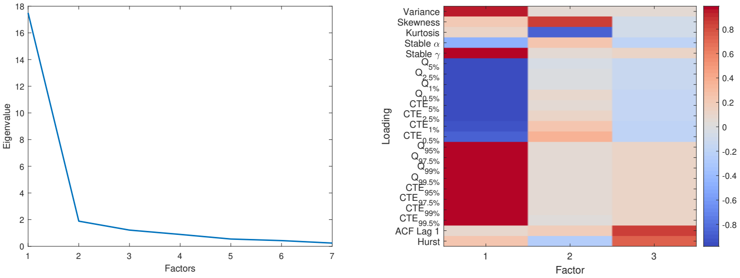


Figure 11: Scree plot and factors loadings. SFA_cryptos

Factor rotation

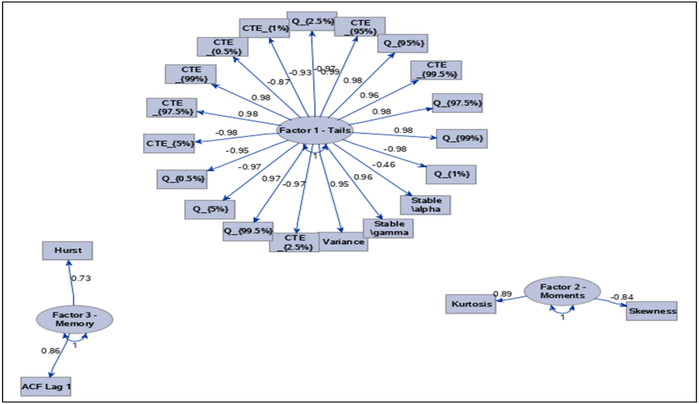


Figure 12: Path diagram. FA_cryptos

Mapping of the factors

1. Tail factor - 76.1% of the total variance
   - Alpha-stable parameters S_α, S_γ
   - Lower and upper quantiles
   - Conditional tail expectations
   - Variance
2. Moment factor - 8.2% of the total variance
   - Skewness
   - Kurtosis
3. Memory factor - 5.3% of the total variance
   - Hurst exponent
   - ACF

Tail factor vs Moment factor

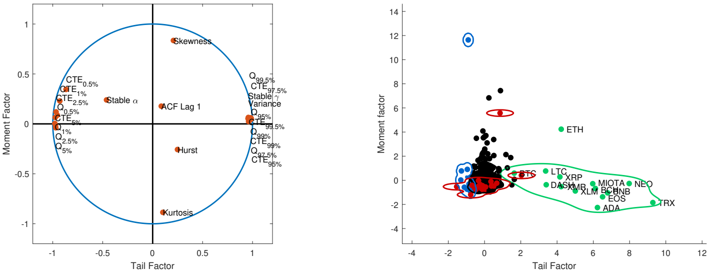


Figure 13: Loadings (left) and scores (right) based on tail and moment factor. SFA_cryptos

Tail factor vs Memory factor

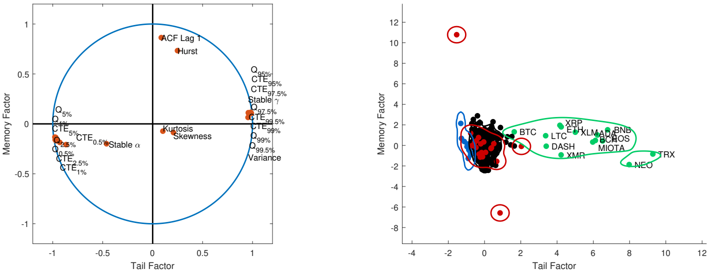


Figure 14: Loadings (left) and scores (right) based on tail and memory factor. SFA_cryptos

Moment factor vs Memory factor

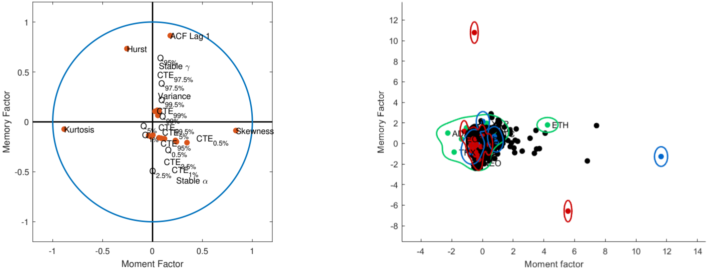


Figure 15: Loadings (left) and scores (right) based on moment and memory factor. SFA_cryptos

Factor explanation

- Classify between Cryptocurrencies and other asset classes
- Binary logistic regression for each factor $F_k,\ k \in \{1,2,3\}$

$$P(Y=1) = \frac{\exp(\beta_0 + \beta_1 F_k)}{1+\exp(\beta_0 + \beta_1 F_k)}, \tag{8}$$

$$Y = \begin{cases} 1, & \text{if Cryptocurrency} \\ 0, & \text{if otherwise} \end{cases} \tag{9}$$

Factor explanation

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

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

$$\tilde{R}^2 = \frac{1 - \left\{ \frac{L(\mathbf{0})}{L(\widehat{\beta})} \right\}^{\frac{2}{n}}}{1 - \{L(\mathbf{0})\}^{\frac{2}{n}}} \tag{10}$$

- $L(0)$ is the likelihood of the intercept-only model
- $L(\widehat{\beta})$ is the likelihood of the full model

Linear Discriminant Analysis

- Finding a projection that maximizes the separability between classes.
- Assumes Gaussianity with equal covariances.

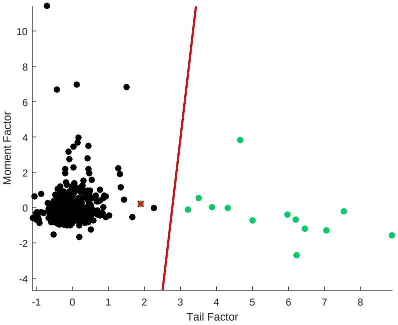


Figure 16: LDA ▸ LDA

Quadratic Discriminant Analysis

- Finding a projection that maximizes the separability between classes.
- Assumes Gaussianity with different covariances.

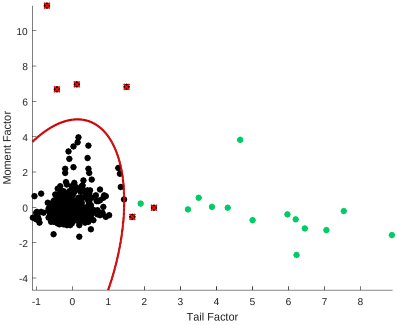


Figure 17: Quadratic Discriminant Analysis

Support Vector Machines

- Finding a projection that maximizes margin in a hyperplane of the original data.
- No parametric assumptions on the underlying probability distribution function.

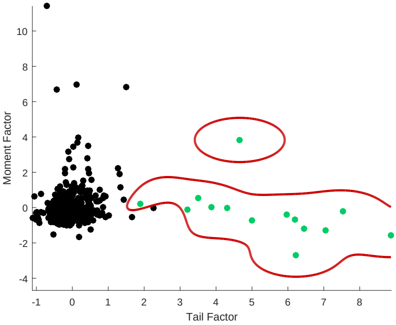


Figure 18: SVM ▸ SVM

Maximum Variance Components Split

- These method have goals to separate, respectively, the components of a structure like the types of assets herein, and clusters defined as the components of a mixture distribution.
- They are based on an unusual variance decomposition in between-group variations.

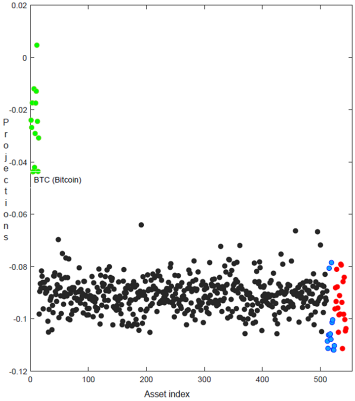
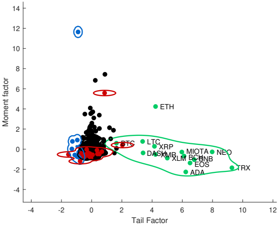


Figure 19: MVCS. VCS_cryptos

MVCS

Video

- Expanding rolling window estimation
  - Starting window 2014-10-22 till 2016-02-20 (1/3 of the data)
  - Increases daily up to full window 2014-10-22 till 2018-10-16
  - Kernel density contour level 0.015
- Clusters converge over time



DFA_cryptos

Phenotypic convergence

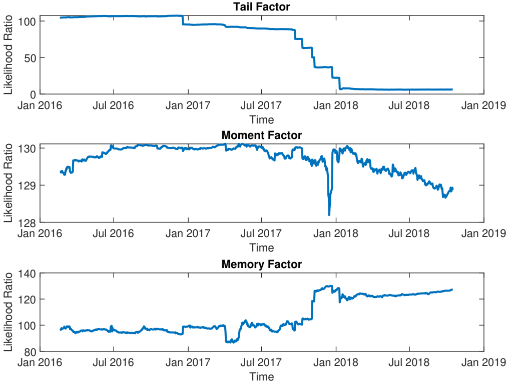


Figure 20: Likelihood Ratios for the binary logistic model, estimated for the period 02/19/2016-10/16/2018. CONV_cryptos

Conclusion

- Financial perspective
  - Main statistical difference between Cryptocurrencies and other asset classes: tail behavior.
  - Moments and memory are of subliminal importance.
  - Nonlinear classification with SVM provides proficient results for risk analysts and regulators.
  - Cryptocurrencies are completely separated by the other types of assets, as proved by Maximum Variance Components Split method.
- Biological perspective
  - Speciation takes time to form distinct species, which potentially evolve further away from each other.
  - Cryptocurrencies establish themselves as unique asset classes.

Exchange rates

▸ Data

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

Cryptocurrencies

Data

1. BTC Bitcoin
2. ETH Ethereum
3. XRP Ripple
4. BCH Bitcoin Cash
5. EOS EOS
6. XLM Stellar
7. LTC Litecoin
8. ADA Cardano
9. XMR Monero
10. TRX TRON
11. BNB Binance Coin
12. MIOTA Iota
13. DASH Dash
14. NEO Neo

Commodities

Data

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

Lévy-Stable distributions

- Fourier transform of characteristic function $\varphi_X(u)$

$$S(X \mid \alpha, \beta, \gamma, \delta) = \frac{1}{2\pi} \int \varphi_X(u) \exp(-iuX) du$$

- Characteristic function representation, $0 < \alpha < 2, \alpha \neq 1$

$$\log \varphi_X(u) = iu\delta - \gamma |u|^{\alpha} \{1 + i\beta (u/|u|) \tan(\alpha\pi/2)\} \tag{11}$$

- Stability or invariance under addition

$$n \log \varphi_X(u) = iu(n\delta) - (n\gamma)|u|^{\alpha} \{1 + i\beta (u/|u|) \tan(\alpha\pi/2)\}$$

- Limiting distribution of n i.i.d. stable r.v., $0 < \alpha \leq 2$
  GCLT (Gnedenko and Kolmogorov, 1954)

$$n^{-\frac{1}{\alpha}} \sum_{i=1}^{n} (X_i - \delta) \xrightarrow{\mathscr{L}} S(\alpha, \beta, \gamma, 0) \tag{12}$$

▸ Statistical assessment

Linear Discriminant Analysis

- Let $X_i \sim N(\mu_i, \Sigma_i)$ belonging to class ω_i, $\Sigma_i = \Sigma_j$
- Project samples X onto a line $Y = w^\top X$
- Select the projection that maximized the separability
- Maximize normalized, squared distance in the means of the classes

$$w^* = \arg\max_w \frac{|w^\top(\mu_i - \mu_j)|^2}{s_i^2 + s_j^2}, \tag{13}$$

$$s_i^2 = \sum_{x_i \in \omega_i} (w^\top x_i - w^\top \mu_i)^2 = w^\top S_i w \tag{14}$$

- Linear Discriminant of Fisher (1936)

$$w^* = S_W^{-1}(\mu_i - \mu_j), \; S_W = S_i + S_j \tag{15}$$

▸ LDA

Support Vector Machines

- Given training data set D with n samples and 2 dimensions

$$D = (X_1, Y_1), \ldots (X_n, Y_n),$$
$$X_i \in \mathbb{R}^2, \quad Y_i \in [0,1]$$

- Finding a hyperplane that maximizes the margin

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

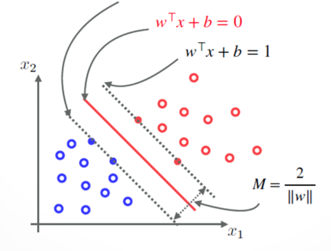


Figure 21: SVM

Variance Component Split

- 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

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \sum_{i=1}^{n-1}\frac{i(n-i)}{n^2}(\bar{X}_{[i+1,n]}-\bar{X}_{[1,i]})(X_{(i+1)}-X_{(i)}). \quad (16)$$

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

$$W_i = W_i(X_1,\ldots,X_n) = \frac{i(n-i)}{n}\frac{(\bar{X}_{[i+1,n]}-\bar{X}_{[1,i]})(X_{(i+1)}-X_{(i)})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} \quad (17)$$

- Index $\mathscr{I}_n = \max\{W_i, i = 1,\ldots,n-1\}$ determines two potential clusters or parts of a structure and is based on averages and inter-point distances.

Maximum Variance Component Split

- 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 $\mathscr{D}_M$ 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 $\mathscr{D}_M$ can be

$$(\Pi_{l=1}^{r} cos\theta_l,\ sin\theta_1 \Pi_{l=2}^{r} cos\theta_l, \ldots,\ sin\theta_{r-1} cos\theta_r,\ sin\theta_r), \tag{18}$$

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

▸ MVCS

Conclusion

- Financial perspective
  - Main statistical difference between Cryptocurrencies and other asset classes: tail behavior.
  - Moments and memory are of subliminal importance.
  - Nonlinear classification with SVM provides proficient results for risk analysts and regulators.
  - Cryptocurrencies are completely separated by the other types of assets, as proved by Maximum Variance Components Split method.
- Biological perspective
  - Speciation takes time to form distinct species, which potentially evolve further away from each other.
  - Cryptocurrencies establish themselves as unique asset classes.