FINTECH RISK MANAGEMENT

www.fintech-ho2020.eu

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

January 27, 2021

# Overview

- Use Case I: Libra or Librae? Basket-based Stablecoins (Paolo Giudici, Thomas Leach and Paolo Pagnottoni, UNIPV)
- Use Case II: ICOs success drivers: a textual and statistical analysis (Paola Cerchiello and Anca Mirela Toma, UNIPV)
- Use Case III: A Statistical Classification 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)
- Use Case IV: Cyber risk management with rank based models and explainable AI (Paolo Giudici and Emanuela Raffinetti, UNIPV)
- Use Case V: Analysis of the cryptocurrency market applying different prototype-based clustering techniques, Luis Lorenzo, Javier Arroyo (Universidad Complutense de Madrid)
- Use Case V: Financial Risk Meter for Cryptos. Rui Ren, Vanessa Guarino, Michael Althof, Anna Shchekina (Humboldt Universität zu Berlin)

# 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.

# 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

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

# Impulse Response Functions and Spillovers

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

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

  - $\Phi_i$ :$(n \times n)$ VAR parameter matrices
  - $k$ : autoregressive order
  - $\varepsilon_t$ : zero-mean white noise process having variance-covariance matrix $\Sigma_\varepsilon$
- 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)

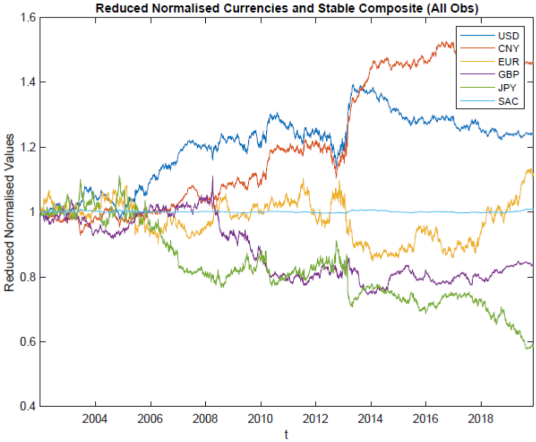



Figure 1: Time evolution of the Reduced Normalised VALue of the basket currencies (USD, CNY, EUR,GBP, JPY), and of the basket based stable coin (SAC)

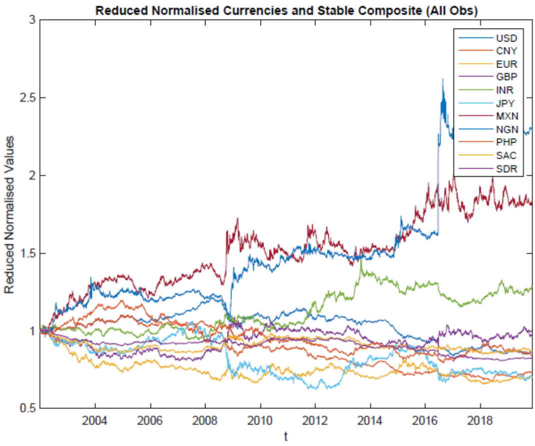



Figure 2: Time evolution of the Reduced Normalised VALue of the basket currencies (USD, CNY, EUR,GBP, JPY), of the considered emerging market currencies (INR, MXN, NGN, PHP) and of the basket based stable coins (SAC, SDR)

| | 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 |
| $\sigma$ | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_{all}$ | 0.09 | 0.14 | 0.07 | 0.06 | 0.13 | 0.11 | 0.22 | 0.41 | 0.10 | 0.04 | 0.05 |
| $\sigma_{pre}$ | 0.05 | 0.02 | 0.04 | 0.03 | 0.04 | 0.07 | 0.11 | 0.35 | 0.03 | 0.01 | 0.03 |
| $\sigma_{cri}$ | 0.02 | 0.04 | 0.03 | 0.03 | 0.12 | 0.07 | 0.05 | 0.04 | 0.03 | 0.02 | 0.02 |
| $\sigma_{post}$ | 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

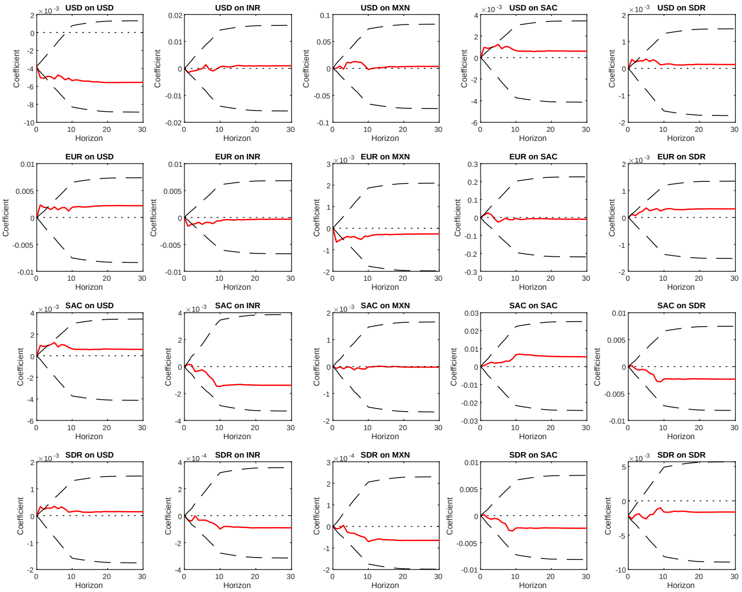



Figure 3: Cumulative impulse response functions

# 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]:

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

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:

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

# 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

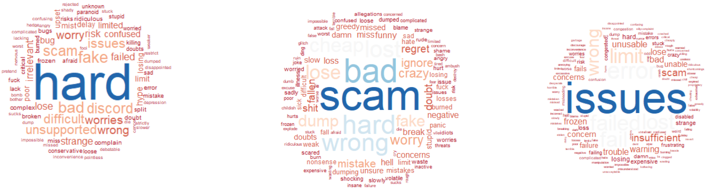



Figure 4: Wordcloud with negative words – failed – scam – success

# 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: A Statistical Classification 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

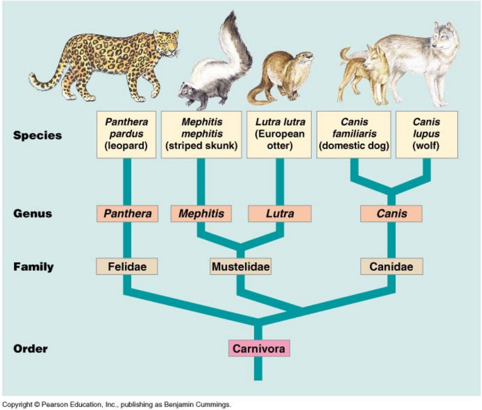



Figure 7: Genus differentia approach in biology

# Genus differentia approach

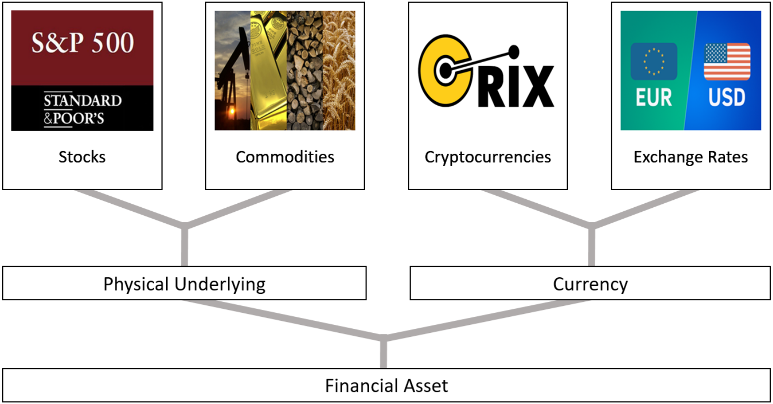


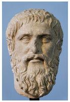
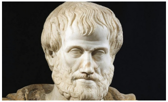

Figure 8: Genus differentia approach in finance

# 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

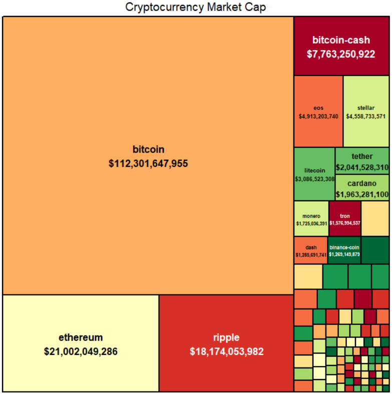



Figure 9: Components of the CRIX cryptocurrency index. Mkt_cryptos

# 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 |
| | $S_\alpha$ | 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

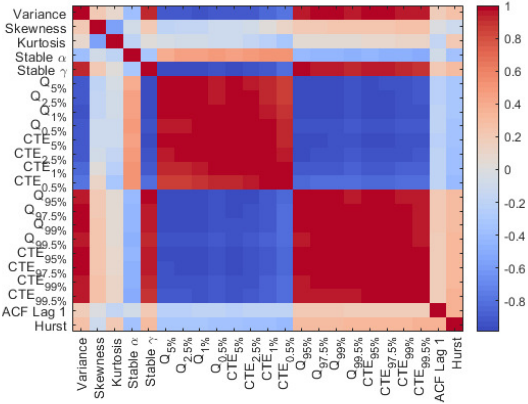



Figure 10: Correlation matrix of the statistical estimates. SFA_cryptos

# 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 ($k < p$)
  - $\mu$ is the vector of the means of initial $p$ variables
  - $\varepsilon$ 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

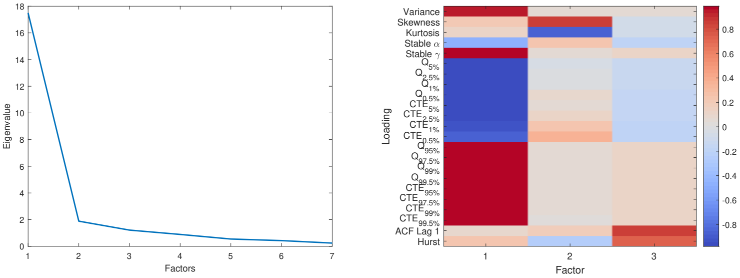



Figure 11: Scree plot and factors loadings. SFA_cryptos

# Factor rotation

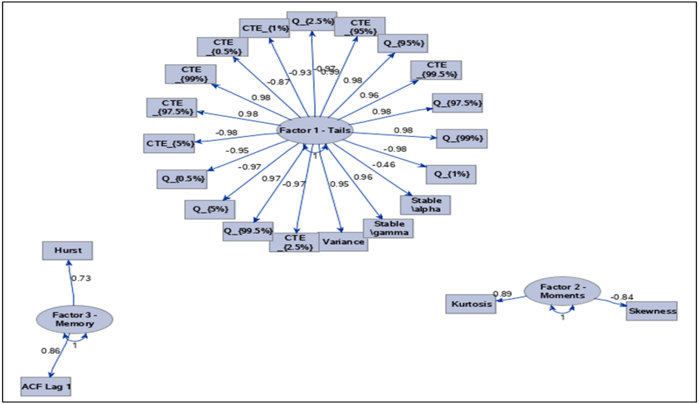



Figure 12: Path diagram. FA_cryptos

# Mapping of the factors

1. Tail factor - 76.1% of the total variance
   - Alpha-stable parameters $S_\alpha$, $S_\gamma$
   - 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

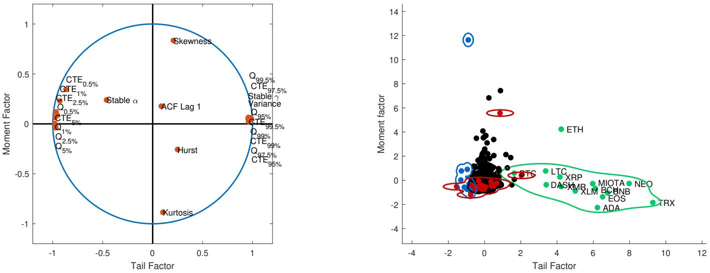



Figure 13: Loadings (left) and scores (right) based on tail and moment factor. SFA_cryptos

# Tail factor vs Memory factor

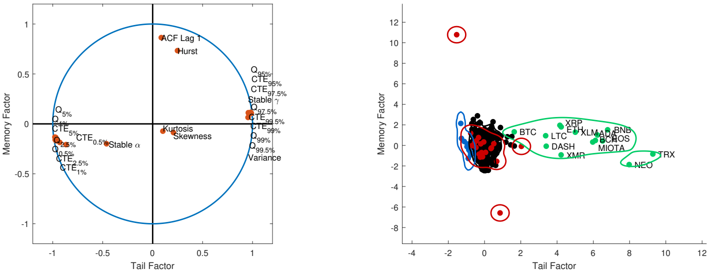



Figure 14: Loadings (left) and scores (right) based on tail and memory factor. SFA_cryptos

# Moment factor vs Memory factor

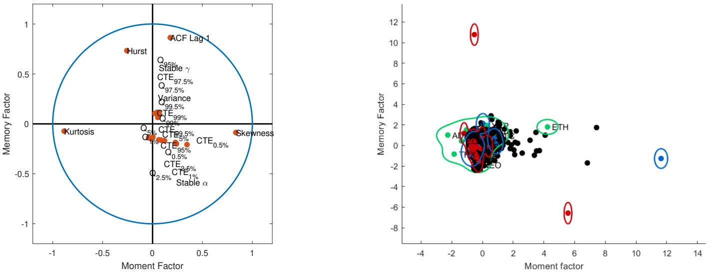



Figure 15: Loadings (left) and scores (right) based on moment and memory factor. SFA_cryptos

# 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 $\beta_1$ | 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.

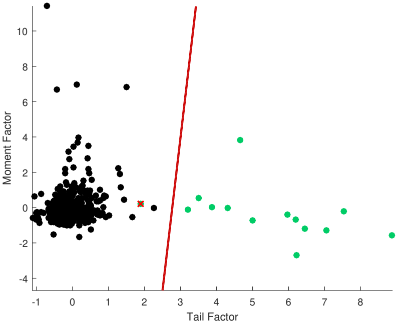



Figure 16: LDA ▸ LDA

# Quadratic Discriminant Analysis

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

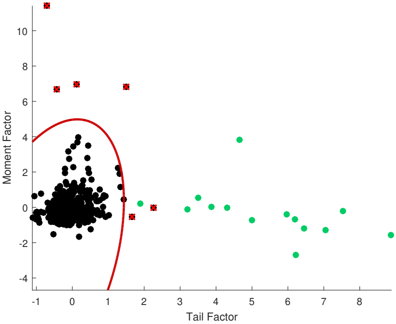



Figure 17: Quadratic Discriminant Analysis

# 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.

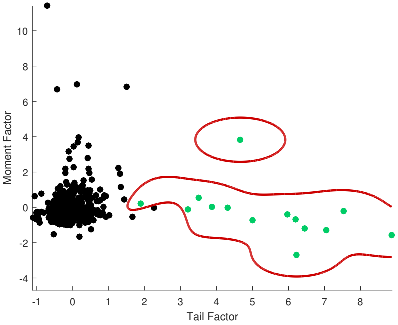



Figure 18: SVM ▸ SVM

# 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.

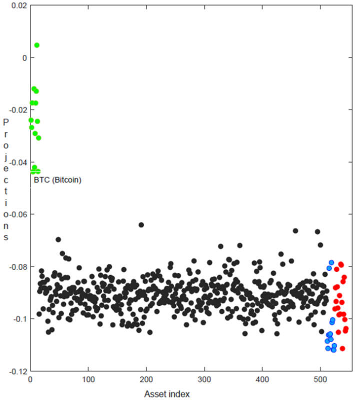
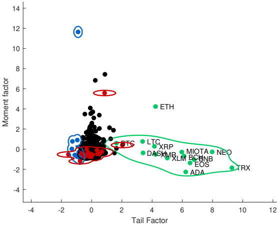



Figure 19: MVCS. VCS_cryptos

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

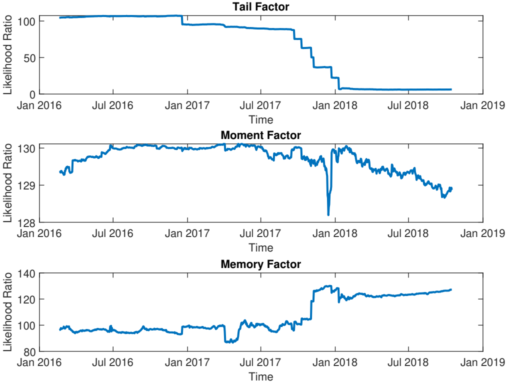



Figure 20: Likelihood Ratios for the binary logistic model, estimated for the period 02/19/2016-10/16/2018. CONV_cryptos

# 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 $\omega_i$, $\Sigma_i = \Sigma_j$
- 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

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

- Finding a hyperplane that maximizes the margin

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

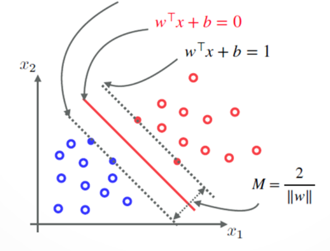



Figure 21: SVM

# 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 $\bar{X}_{[i+1,n]}$, $i = 1,\ldots,n-1$, 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 $X_{(1)},\ldots,X_{(i)}$ and $X_{(i+1)},\ldots,X_{(n)}$ 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 **a** coincides with a component of the structure we have complete separation of this component along **a**.
- A set of projection directions $\mathscr{D}_M$ can be

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

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

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.

# Use Case IV: Cyber risk management with rank based models and explainable AI (Paolo Giudici and Emanuela Raffinetti, UNIPV)

# Cyber risk: definition and main concern

## Cyber risk:

"any risk emerging from the use of information and communication technology (ICT) that compromises the confidentiality, availability, or the integrity of data or services".

Cybersecurity has become a serious concern for businesses, among operational risks.

## Operational risk:

"the risk of a monetary loss caused by human resources, IT systems, by organisation processes or by external events".

While the literature on the quantitative measurement of operational risks, based on loss data, constitutes a reasonably large body, that on cyber risk measurement is very limited. This may be due to the limited availability of data which are typically not disclosed.

## Cyber risk model specification - Rank Regression Model

Let $Y$ be a response variable, expressed through $k$ ordered categories (severity of cyber attacks).

### Procedure:

- assign a rank $r_1 = 1$ to the smallest ordered category of $Y$;
- assign rank $(r_{j-1} + n_{j-1})$ to the following ordered categories, where $n_{j-1}$ is the absolute frequency associated with the $(j-1)$-*th* category with $j = 2, \ldots, k$;
- the phenomenon described by the $Y$ variable can be re-formulated in terms of its ranks $R$, where:

$$R = \left\{ \underbrace{r_1, \ldots, r_1}_{n_1}, \underbrace{r_2, \ldots, r_2}_{n_2}, \ldots, \underbrace{r_k, \ldots, r_k}_{n_k} \right\},$$

  with $r_1 = 1$, $r_2 = r_1 + n_1$ and $r_k = r_{k-1} + n_{k-1}$.
- Given $p$ explanatory variables, a regression model for $R$ can be specified as:

$$\hat{R} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_p X_p,$$

  whose unknown parameters can be estimated by the OLS method.

# Cyber risk model specification - Model validation

To validate the model, we resort to a recently developed criterion called *Rank Graduation Accuracy* ($RGA$), which is based on the Lorenz ($L_Y$), dual Lorenz ($L_Y^{'}$) and concordance ($C$) curves.

Being $Y = R$, the $RGA$ index is defined as:

$$RGA = \sum_{i=1}^{n} \frac{\left\{(1/(n\bar{r}))\sum_{j=1}^{i} r_{ord(\hat{r}_j)} - i/n\right\}^2}{i/n} = \sum_{i=1}^{n} \frac{\left\{C(r_{ord(\hat{r}_j)}) - i/n\right\}^2}{i/n}$$

where $r_{ord(\hat{r}_j)}$ are the rank-transformed response variable values, re-ordered by the ranks predicted by the model, $\bar{r}$ is the mean of all ranks and $C(r_{ord(\hat{r}_j)}) = \frac{\sum_{j=1}^{i} r_{ord(\hat{r}_j)}}{\sum_{i=1}^{n} r_{ord(r_i)}}$.
$RGA_{norm}$ can be specified as the ratio between $RGA$ and its maximum value ($RGA_{norm} \in [0,1]$).

# Cyber risk model evaluation - A novel statistical test

The test statistics is formalized as:

$$T = \sum_{i=1}^{n} \frac{\left\{ \sum_{j=1}^{i} R_{ord(\hat{r}_j)} - \sum_{j=1}^{i} R^{e}_{ord(\hat{r}_j)} \right\}^2}{\sum_{j=1}^{i} R^{e}_{ord(\hat{r}_j)}},$$

where $R^{e}_{ord(\hat{r}_j)}$ is the expected rank-transformed response variable value in the case of a random model and $T \sim \chi^2_n$.

The comparison may occur between a full model (including all the covariates in the dataset) and a reduced model (including only some of the covariates in the dataset).

- Define with $T_{full}$ the test statistics $T$ computed under the full model.
- Define with $T_{reduced}$ the same statistics computed under the reduced model.
- Let $T_{model} = T_{full} - T_{reduced}$ be the difference between the two test statistics.
- The following proposition holds:

***Proposition.*** $T_{model} = T_{full} - T_{reduced}$ is distributed as a Variance Gamma distribution, with parameters $\lambda = n/2$, $\alpha = 1/2$, $\beta = 0$ and $\mu = 0$, where $\lambda$ and $\alpha \in \mathbb{R}$, $\beta$ is the asymmetry parameter and $\mu$ is the location parameter.

## Cyber risk model implementation - A remark

If we resort to the VarianceGamma R package, the parameter $\lambda$ must be set equal to one half the number of observations included in the dataset. But the computation of the $p$-values of the VarianceGamma associated with the R package is not possible when $\lambda$ takes large values.

A possible solution is to focus only on samples of small size which may be directly drawn by the dataset.

To overcome this problem we refer to the subsampling procedure introduced by Raffinetti and Romeo (2015).

- Consider a number $h$ of different samples and compute for each sample the value of the test $T_{model}$.
- The Variance Gamma distribution is symmetric around zero. This consideration, suggests the use of a bilateral test.

## Cyber risk model implementation - The *s*-value and the *s*-scale

A *significance value* (named *s*-value) defined as the relative percentage of significant tests

$$s\text{-value} = P(T_{imodel} \geq |t_{\alpha/2}|) = \frac{1}{h}\sum_{i=1}^{h} I_{T_{imodel} \geq |t_{\alpha/\mathbf{2}}|}, \quad i = 1, \ldots, h,$$

where

$$I_{T_{imodel} \geq |t_{\alpha/\mathbf{2}}|} = \begin{cases} 0, \text{ if } -t_{\alpha/2} < T_{imodel} < t_{\alpha/2} \\ 1, \text{ otherwise.} \end{cases}$$

is employed.

The *s*-value can be associated with a significance scale (*s*-scale) summarised as follows:

| *s*-value classes | *s*-classes levels |
|---|---|
| *s*-value=1 | Always significant |
| 0.7<*s*-value<1 | Almost always significant |
| 0.5<*s*-value≤0.7 | Frequently significant |
| 0.3<*s*-value≤0.5 | Sometimes significant |
| 0<*s*-value≤0.3 | Rarely significant |
| *s*-value=0 | Never significant |

# Data

Our proposal is applied to real loss data, organised by severeity levels, reported in the italian annual report on cyber risks (Clusit, 2018).

We focus on a sample data, consisting of 808 cyber attacks observed in 2017.

Severity levels are reported according to the type and technique of attacks (which can be seen as event types), the victims and their country of origin (which can be seen as business lines).

# Analysis

The aim is to detect the main factors which may affect the severity degree.
We consider three rank regression models which differ in terms of the variables taken into account:

- the first rank regression model is built on all the explanatory variables appearing in our dataset (cyber attacks, attack techniques, victim type and continent);
- the second rank regression model was specified by removing from the full model the continent variable, in order to assess if the geographical area where the cyber attacks occur may impact on the severity degree;
- the third rank regression model was built by removing from the full model the cyber attack type variable.

# Results

| | Full model | | Reduced model (without continent variable) | |
|---|---|---|---|---|
| *Coefficient* | *Estimate* | *p-value* | *Estimate* | *p-value* |
| Intercept | 87.42 | 0.02678 | 175.65 | 0.01615 |
| Espionage/Sabotage | -231.38 | <0.001 | -231.88 | <0.001 |
| Hacktivism | -39.210 | 0.00663 | -38.99 | 0.00672 |
| Information warfare | -222.17 | <0.001 | -221.71 | <0.001 |
| Entertainment/News | 117.14 | 0.03345 | 115.53 | 0.03549 |
| GDO/Retail | 139.97 | 0.01743 | 138.18 | 0.01855 |
| Online Services/Cloud | 136.11 | 0.01496 | 135.52 | 0.01514 |
| Research-Education | 142.26 | 0.01057 | 140.07 | 0.01158 |
| Phishing/Social Engineering | 120.27 | 0.01763 | 120.63 | 0.01708 |
| Unknown | 99.670 | 0.04516 | 100.21 | 0.04357 |

**Categorical variable reference level:** cyber attack (first block): Cybercrime; victim type (second block): Automotive; attack technique (third block): 0-day

Remark:
In the reduced rank regression model **without the cyber attack variable**, Entertainment/News is no more significant while DDoS, Malware, Malware and Vulnerabilities become significant.

| **Model** | *RGA* | $RGA_{norm}$ | RMSE |
|---|---|---|---|
| Full rank regression model | 63.185 | 0.739 | 105.196 |
| Reduced rank regression model (without continent variable) | 63.111 | 0.738 | 105.284 |
| Reduced rank regression model (without cyber attack variable) | 47.426 | 0.555 | 122.706 |

# Testing the additional contribution

The assessment of the significance in the difference between the full model and the reduced model without the cyber attack is led by resorting to the subsampling procedure proposed by Raffinetti and Romeo (2015).

We specify:

- $\alpha = 0.05$, as significance level;
- n=10, as the sample size;
- $h = \{100, 500, 1,000\}$, as the number of samples drawn from the dataset.

| $h$ | $s$-value | $s$-scale |
|---|---|---|
| 100 | 0.750 | Almost always significant |
| 500 | 0.788 | Almost always significant |
| 1,000 | 0.811 | Almost always significant |

# The Shapley-Lorenz decomposition

The Shapley-Lorenz decomposition appears as new eXplainable Artificial Intelligence method and can be expressed as:

$$LZ^{X_k}_{d=1}(\hat{Y}) = \sum_{X^{'} \subseteq \mathscr{C}(X) \setminus X_K} \frac{|X^{'}|!(K - |X^{'}| - 1)!}{K!} [LZ_{d=1}(\hat{Y}_{X^{'} \cup X_k}) - LZ_{d=1}(\hat{Y}_{X^{'}})],$$

where $LZ_{d=1}(\hat{Y}_{X^{'} \cup X_k})$ and $LZ_{d=1}(\hat{Y}_{X^{'}})$ describe the (mutual) variability explained by the models including the $X^{'} \cup X_k$ variables and the $X^{'}$ variables, respectively.

## Results

| Additional covariate ($X_k$) | $LZ^{X_k}_{d=1}(\widehat{Severity})$ | Global Shapley |
|---|---|---|
| Type of attack | 0.072 | -748.96 |
| Type of victim | 0.115 | 5.15 |
| Technique of attack | 0.058 | -34.36 |
| Continent | 0.032 | -25.67 |

# Summary

- We have proposed a novel statistical model to measure cyber risks, which can be applied to the observed levels of severity.
- Our proposal can be employed as an ordinal effective measurement to prioritise cyber risk.
- Its application to a real cyber loss database, measured at the ordinal level, reveals that the proposed model tools are indeed able to individuate and explain the most important types of cyber attacks.
- The proposed model could be the basis to investigate dependency structures among risks, that reveal forms of contagion.

# Use Case V: Analysis of the cryptocurrency market applying different prototype-based clustering techniques, Luis Lorenzo, Javier Arroyo (Universidad Complutense de Madrid)

# Motivation

What?

- Since the appearance of the Bitcoin in 2008, the electronic market of cryptocurrencies has grown up to more than 4,000 cryptocurrencies with over 800 trades per second in more than 280 exchanges.
- Market with a high heterogeneity:
  - Blockchain technology with a huge variety of encryption and consensus algorithms
  - High diversity on the cryptoasset's usage: cryptocurrency - cryptocoin – criptocomodities - criptotokens

# Goals

- *Big Why*: general goals
- Discover if there are underlying structures in the market
- Summarize and segment the whole cryptocurrency market in 2018
- Point out the relevant variables of the market that could help us to characterize the groups of the cryptocurrencies

- *Small Why*: other specific targets
- Supported on specific data mining methodology, we want to reach a fine investment recommendation for particular cryptocurrencies
- Foreseen of potential target for a research, mainly on portfolio theory and that could help in a more efficient allocation of assets.

# Methods

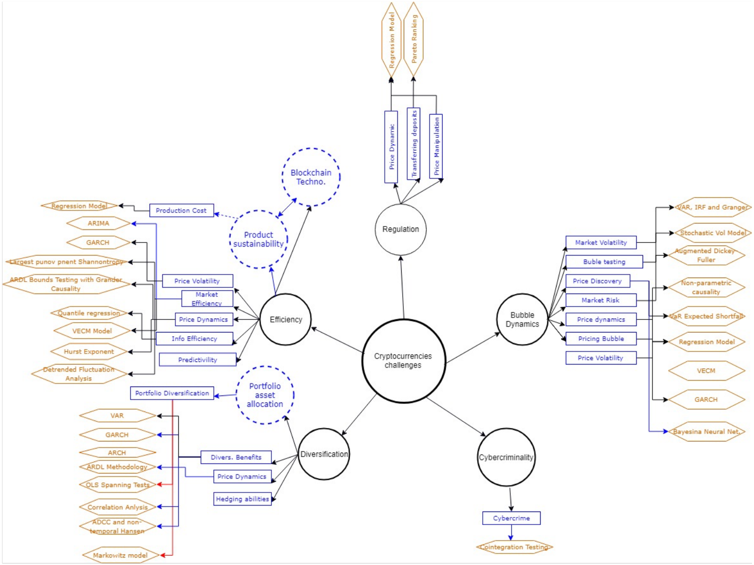
How?

- Supported in data mining techniques
- Inspired on what have been done on traditional financial markets
- Axes:
  - Data mining techniques mainly supported on different clustering techniques and association tests
  - Taken some performance indicator from Portfolio Optimization theory
  - Understanding of the specificities of the cryptoasset market due to the underlying blockchain technology

# Hints

- Some hints on the methodology (how)
- Prototype-based clustering methods will help for a good representation of the market summarizing the trends on the market and simplifying the dimensionality of the market
- Taken benefit of different levels (dimensions) of the clustering objects (granularity):
  - Mean and standard deviation $(\mu, \sigma)$
  - Distribution –Histogram
  - Time-series
- Intersection of the clusters: we combine all the clusters
- In all cases, we apply association tests between the cryptocurrencies allocated in the clusters and some categorical variables of the market (technological and finance ratios)

# Cryptocurrencies research-landscape on papers



# Methodology

- Dataset preparation
  - All cryptocurrencies traded the 95% of the time along 2018
  - 10,131,981 observations, 8 variables (time, open, close, high, low, volumefrom, volumeto, sym):
- Final dataset: 630,618 observations for 1,723 cryptocurrencies
- Variables:
  - Daily log-return:
  - Market cap
  - Technological variables:
    - Encryption algorithm
    - Consensus algorithm
  - Age: maturity
  - Financial ratios

# Methodology

Data-preparation

- Financial ratios (asset portfolio KPIs) :
  - $\beta = \frac{Cov(R_c.R_b)}{Var(R_b}$
  - Sharpe-ratio
    - $SR_c = \frac{E[R_c - R_f]}{\sigma_c}$
    - Modigliani-Miller: $MM_c = SR_c \sigma_b + E[R_f]$
  - Sortino
    - $Sort_c = \frac{E[R_c - R_T]}{\sigma_D}$
    - $M_S^2 = R_c + Sort_c(\sigma_{Db} - \sigma_D)$
  - Downside-risk: $\sigma_D = \sqrt{\sum_{i=1}^{n} \frac{min(R_c - R_T, 0)^2}{n}}$

# Methodology

We aim to find groups of cryptocurrencies by clustering techniques on the log-returns along 2018

- Bi-Dimensional Clustering (K-means)
  - Golden rule for clustering and acts as our benchmark
  - Variables are standardized
  - The number of clusters chosen by Cluster Validity Index and squared Euclidean Distance
  - Ensemble of repeated runs (500) to find the more representative partition.
- Dynamic clustering of histogram (Hist DAWass algorithm)
  - Clustering algorithm on histogram data based on the $L_2$ Wasserstein distance
  - $d_w(h_1, h_2) = \sqrt{\int_0^1 [F1^-1(t) - F2^-1]^2 dt}$
  - Quality measure based on sum of squares deviation of the model
  - Repeated run 20 times

# Methodology

- Clustering of time-series (TADpole algorithm)
  - Adaptation of Dynamic Time Warping with a faster algorithm
  - Deterministic method that only require 2 parameters:
  - Cut-off distance (2); threshold to select series
  - Window-size (3); time-frame to make the comparison
- Intersection of clusters
  - We could have $T_1 \times T_2 \times T_3 = 45$ intersections but only 23 are populated, however we keep only the higher cardinality ones (6 intersections in total)

| Intersection | Kmeans | HistDAWass | TADPole | N |
|---|---|---|---|---|
| 1 | 3 | 3 | 3 | 295 |
| 2 | 3 | 3 | 2 | 294 |
| 3 | 1 | 3 | 3 | 208 |
| 4 | 1 | 3 | 2 | 196 |
| 5 | 1 | 2 | 3 | 166 |
| 6 | 1 | 2 | 2 | 148 |

# Methodology

Association tests

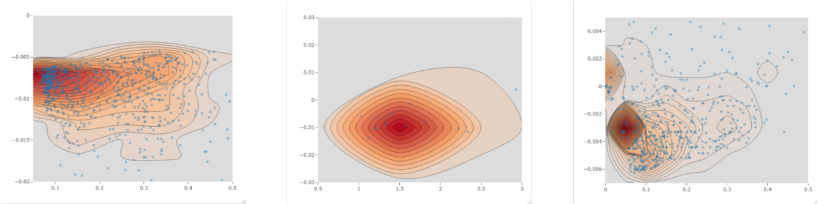
- Typically, Chi-Square test is used to examine the significance of the association between categorical data on a contingency table.
- But we have some constraints with some frequencies for some combinations of variables (not >5 in at least 80% of the cells)
- We apply Fisher exact tests applicable for all sample sizes
- We run 8,000 Montecarlo simulations for each association, assuming a multivariate hypergeometric distribution in the contingency tables
- For each significant association between categorical variables on the contingency table, we analyze the Standardized Pearson's Residual:

$$r_{(adj)ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}(1 - \frac{m_i}{N})(1 - \frac{n_j}{N})}}$$

# Results

Low number of clusters for all techniques, with two groups clearly dominant in K-means and TADPole

| | **K-means** | | | **HistDAWass** | | | **TADPole** | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Card.* | *Mean* | *S.D.* | *Card.* | *Mean* | *S.D.* | *Card.* | *Mean* | *S.D.* |
| **Clus. 1** | 801 | -0.009 | 0.229 | 147 | -0.503 | 0.378 | 22 | -0.001 | 0.080 |
| **Clus. 2** | 19 | -0.008 | 1.795 | 496 | -0.134 | 0.337 | 843 | 0.026 | 0.046 |
| **Clus. 3** | 903 | -0.002 | 0.130 | 1007 | -0.011 | 0.108 | 858 | -0.028 | 0.047 |
| **Clus. 4** | | | | 57 | -0.044 | 0.867 | | | |
| **Clus. 5** | | | | 16 | -0.095 | 3.123 | | | |



(a) Cluster 1 (b) Cluster 2 (c) Cluster 3

Figure 22: High Heterogeneity in Cluster 3

# Results

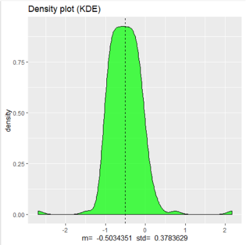
DAWass measures

- Coefficient of Variation makes the difference
- Clusters highly homogeneous based on the variance for Wasserstein distance

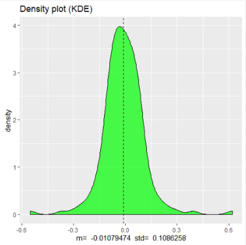
| | Mean | Std. Dev. | Coef.Var. | Skew. | Kurt. | Med. | Min. | Max. | Var.Wass. |
|---|---|---|---|---|---|---|---|---|---|
| Clus. 1 | -0.50 | 0.38 | -0.75 | 0.56 | 9.33 | -0.51 | -2.69 | 2.18 | 0.079 |
| Clus. 2 | -0.13 | 0.34 | -2.51 | 0.82 | 13.43 | -0.16 | -2.24 | 2.36 | 0.025 |
| Clus. 3 | -0.01 | 0.11 | -10.06 | 0.28 | 7.10 | -0.01 | -0.55 | 0.62 | 0.005 |
| Clus. 4 | -0.04 | 0.87 | -19.97 | 0.54 | 11.95 | -0.08 | -5.44 | 6.67 | 0.128 |
| Clus. 5 | -0.09 | 3.12 | -32.90 | 0.05 | 5.66 | -0.17 | -17.56 | 17.56 | 1.116 |

# Results

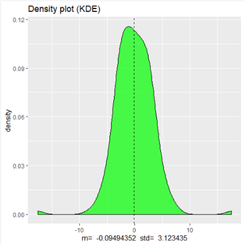
## DAWass measures




(a) Prototype 1




(b) Prototype 2

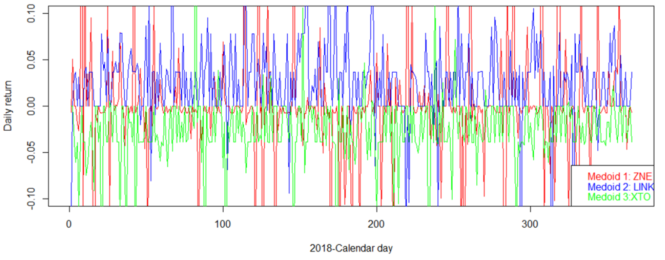



(c) Prototype 3

Figure 23: Prototypes of the DAWass clustering

# Results

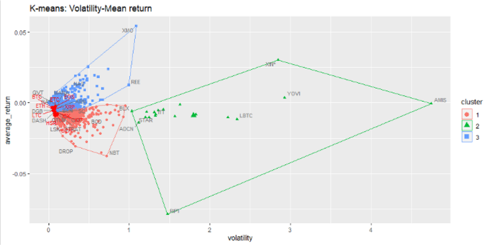
Clear different trends for the medoids: ZNE more positives peaks, XTO more negative peaks and LINK on-average around zero

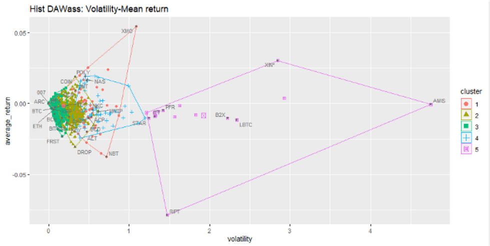



| Cluster | Mean Dist. | Std. Dev. | Coef. Var. |
|---|---|---|---|
| 1 | 4.31 | 3.04 | 0.71 |
| 2 | 4.60 | 3.29 | 0.72 |
| 3 | 4.85 | 3.53 | 0.73 |

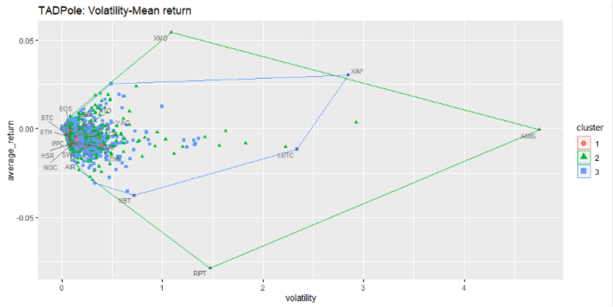
# Results

Volatility-Mean return plane (comparing the outcomes of the different clustering techniques in the same dimensionality plane)



# Results

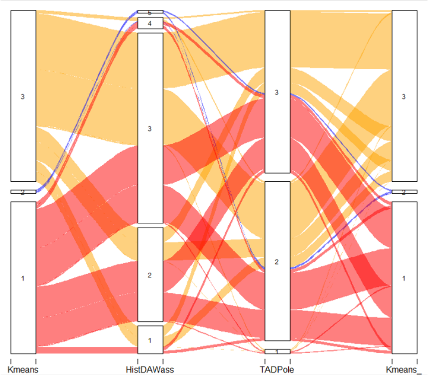



- High overlapping for the objects grouped by the different techniques ( different object dimensionality)
- Exception: Kmeans-Cluster 2 and Hist DAWass-Cluster 5

# Results

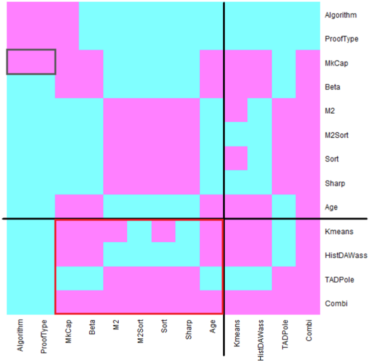
Flow of the cryptocurrencies through the three clustering techniques

- Cluster 3 in K-means and Hist DAWass share a high percentage of cryptocurrencies
- Odd clusters 2 and 5 in K-means and Hist DAWass share the cryptocurrencies
- Divergence of the main stream between Hist DAWass and TADPole



# Results

- Association between MkCap, Beta and Modigliani-Modiglini ration and K-means clusters
- Association between most of the financial ratios and TADPole clustering
- Hist DAWass seems more associated with MkCap, Beta and Age (maturity)
- Technological variables seems associated with MkCap



# Results

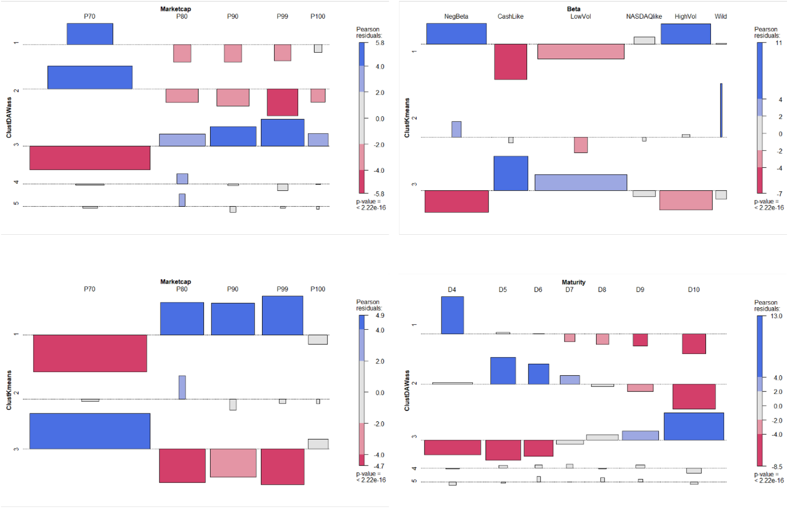



Figure 24: Pearson's residual representation for more significant associations (Beta, Maturity)

# Results

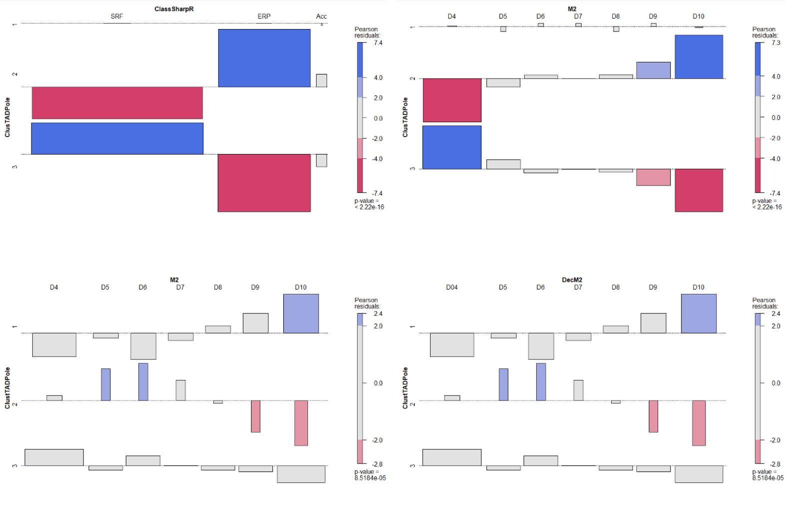



Figure 25: Pearson's residual representation for more significant associations (financial ratios)

# Results

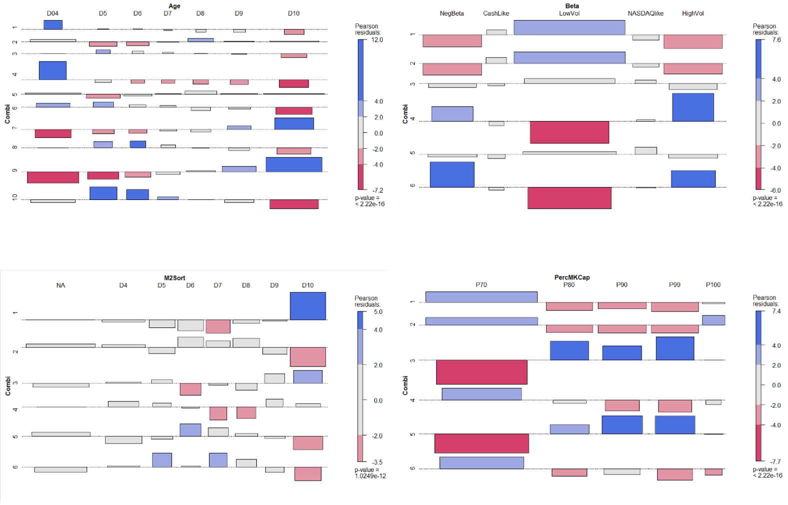



Figure 26: Pearson's residual representation for more significant associations (Intersections)

# Conclusions

Conclusions

- We confirm the existence of structures underlaying the market
- K-means respond quite well to separate the cryptocurrencies in positive, negative returns and high volatility cryptocurrencies
- Hist DAWass algorithms seems more sensitive to market cap and the maturity screening than other techniques
- K-means works quite well to screen by Beta but does not distinguish between high positive and negatives
- TADPole is quite accurate to screen by the financial ratios, i.e. the Sharpe ratio, Sortino and Modigliani-Modigliani
- The intersection of clusters inherit the associations detected in the original but the significance of the association is lower

Next steps:

- Our findings could help to enhance the performance of predictive algorithms
- The investment portfolios can take benefits enhancing the allocation of cryptoassets

# Use Case VI: Financial Risk Meter for Cryptos.

Rui Ren, Vanessa Guarino, Michael Althof, Anna Shchekina (Humboldt Universität zu Berlin)

# FRM financialriskmeter for Cryptos

Michael Althof

Vanessa Guarino

Rui Ren

Anna Shchekina

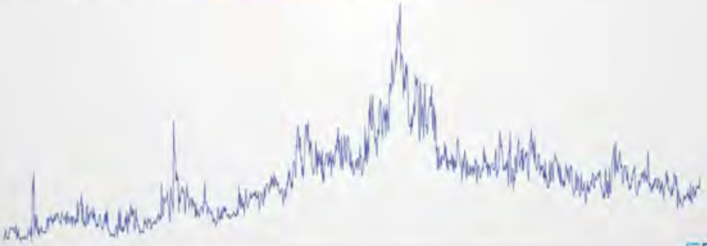
Humboldt-Universität zu Berlin
lvb.wiwi.hu-berlin.de
Charles University, WISE XMU, NCTU 玉山学者

## Tail Events (TE)

- TEs across Cryptos indicate increased risk
- CoVaR measures joint TEs between 2 risk factors
- CoVaR and other risk factors?
- TENET Tail Event NETwork risk, Härdle Wang Yu (2017) J E'trics
- FRM Financial Risk Meter for joint TEs

## Risk Measures

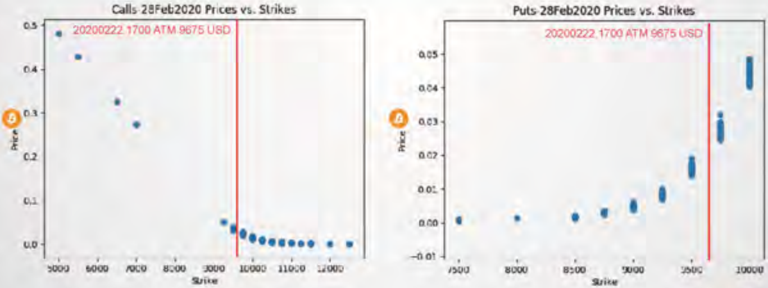
- VIX: IV based, does not reflect joint TEs
- CoVaR concentrates on a pair of risk factors
- CISS, Google trends, SRISK, ...
- FRM displays the full picture of TE dependencies
- Firamis.de/FRM financialriskmeter

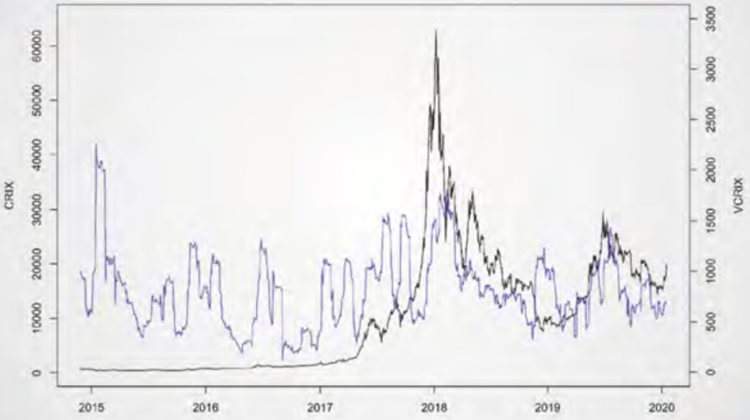


# Call and Puts on BTCs

- Listed at Bloomberg since 20200113

Prices from 20200221 1600 - 20200222 1100
Timestamps precise in the range 1E-3 sec
Calls, Puts with maturity 20200228




CRIX
VCRIX
2015
2016
2017
2018
2019
2020
0
10000
20000
30000
40000
50000
60000
500
1000
1500
2000
2500
3000
3500

# Outline

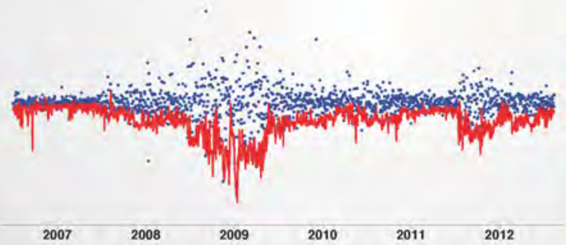
1. Motivation ✔
2. Genesis
3. Framework
4. Applications
5. Node influence metrics
6. Sensitivity analysis
7. Network centrality
8. Portfolio Construction
9. Conclusions

## VaR Value at Risk

- Probability measure based

$$P(X_{i,t} \leq VaR^{\tau}_{i,t}) \stackrel{\text{def}}{=} \tau, \quad \tau \in (0,1)$$

- $X_{i,t}$ log return of risk factor (institution) $i$ at $t$
- VaRs (0.99, 0.01) based on RMA, Delta Normal Method




# Quantiles and Expectiles

For r.v. $Y$ obtain tail event measure:

log returns

$$q^\tau = \arg\min_\theta \mathrm{E}\left\{\rho_\tau(Y-\theta)\right\}$$

asymmetric loss function

$$\rho_\tau(u) = |u|^c\,|\tau - \mathrm{I}_{\{u<0\}}|$$

$c = 1 \succ$ quantiles

$c = 2 \succ$ expectiles

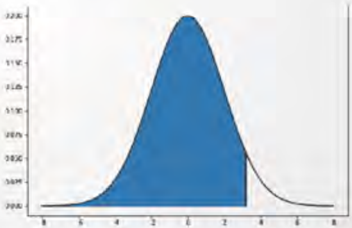


Figure: Quantile of $N(0,2)$, $\tau = 0.7$, $q^\tau = 3.2$

▶ Expectile as Quantile

## Conditional Value at Risk

- Adrian and Brunnermeier (2016) introduced CoVaR
$$P\{X_{j,t} \leq CoVaR^{\tau}_{j|i,t} \mid X_{i,t} = VaR^{\tau}(X_{i,t}), M_{t-1}\} \stackrel{\text{def}}{=} \tau$$
- $M_{t-1}$ vector of macro-related variables
- Nonlinear features, $\tau = 0.05$

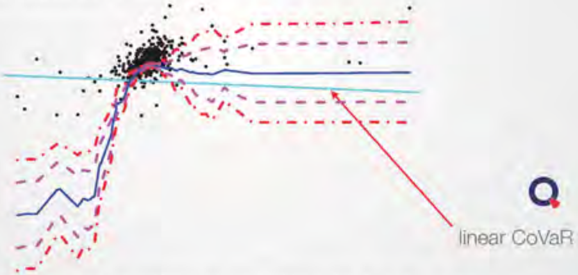



Figure: Goldman Sachs (Y), Citigroup (X), Confidence Bands, see Chao et al (2015)

# CoVaR and the magic of joint TEs

- CoVaR technique

$$X_{i,t} = \alpha_i + \gamma_i^\top M_{t-1} + \varepsilon_{i,t}$$

$$X_{j,t} = \alpha_{j|i} + \beta_{j|i} X_{i,t} + \gamma_{j|i}^\top M_{t-1} + \varepsilon_{j,t}$$

- $F^{-1}_{\varepsilon_{i,t}}(\tau | M_{t-1}) = 0$ and $F^{-1}_{\varepsilon_{j,t}}(\tau | M_{t-1}, X_{i,t}) = 0$

$$\widehat{VaR}^{\tau}_{i,t} = \widehat{\alpha}_i + \widehat{\gamma}_i^\top M_{t-1}$$

$$\widehat{CoVaR}^{\tau}_{j|i,t} = \widehat{\alpha}_{j|i} + \widehat{\beta}_{j|i} \widehat{VaR}^{\tau}_{i,t} + \widehat{\gamma}_{j|i}^\top M_{t-1}$$

CoVaR: First calculate VaRs, then compute the TE given a stressed risk factor.

# Linear Quantile Lasso Regression

$$r_{j,t}^s = \alpha_{j,t}^s + A_{j,t}^{s\top} \beta_j^s + \varepsilon_{j,t}^s \tag{1}$$

$$A_{j,t}^{s\top} \stackrel{\text{def}}{=} \left[ M_{t-1}^s, r_{-j,t}^s \right]$$

where:

- $r_{-j,t}^s$ log returns of all cryptos except $j \in 1:J$ at $t \in 2:T$
- $s$ length of moving window
- $M_{t-1}^s$ log return of macro prudential variable at time $t-1$
- For application, consider $J = 15$, $s = 63$

▶ Crypto List ▶ Macroprudential

## Lasso Quantile Regression

$$\min_{\alpha_j^s, \beta_j^s} \left\{ n^{-1} \sum_{t=s}^{s+(n-1)} \rho_\tau (r_{j,t}^s - \alpha_j^s - A_{j,t}^{s\top} \beta_j^s) + \lambda_j^s \parallel \beta_j^s \parallel_1 \right\} \qquad (2)$$

- Check function $\rho_\tau(u) = |u|^c \, |\tau - \mathrm{I}_{\{u<0\}}|$ with $c = 1, 2$ corresponding to quantile, expectile regression
  - $\lambda$ creates size of „active set", i.e. spillover
  - $\lambda$ is sensitive to residual size, i.e. TE size
  - $\lambda$ reacts to singularity issues, i.e. joint TEs

# $\lambda$ Role in Linear Lasso Regression

- Osborne et al. (2000)
- Dependence, time-varying, institution-specific
- Size of model coefficients depends on,

$$\lambda = \frac{(Y - X\beta(\lambda))^{\top} X\beta(\lambda)}{\| \beta \|_1}$$

Coeff's depend on $\lambda$

- $\lambda$ depends on:
  - Residual size
  - Condition of design matrix
  - Active set

## $\lambda$ Role in Linear Quantile Regression

- $\lambda$ size of estimated LQR coefficients Li Y, Zhu JL (2008)

$$\lambda = \frac{(\alpha - \gamma)^{\top} X\beta(\lambda)}{\|\beta\|_1}$$

Coeff's ( $\lambda$ )

$$(\alpha - \gamma)^{\top} = \tau\, \mathrm{I}_{\{Y - X\beta(\lambda) > 0\}} + (\tau - 1)\, \mathrm{I}_{\{Y - X\beta(\lambda) < 0\}}$$

- Average penalty: indicator for tail risk,

$$FRM^t \stackrel{def}{=} J^{-1} \sum_{j=1}^{J} \lambda_j^t$$

- The FRM time series is one index for joint TEs!

## $\lambda$ Selection

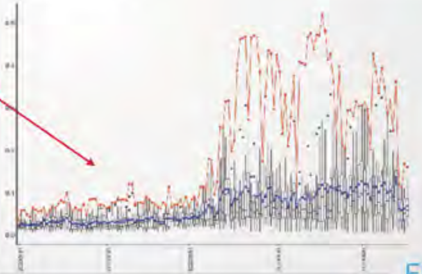
- Generalized approximate cross-validation (GACV) (Yuan, 2006)

$$\min GACV(\lambda_j^s) = \min \frac{\sum_{t=s}^{s+(n-1)} \rho_\tau(r_{j,t}^s - \alpha_j^s - A_{j,t}^{s\top}\beta_j^s)}{n - df} \quad (3)$$

Coeff's depend on $\lambda$

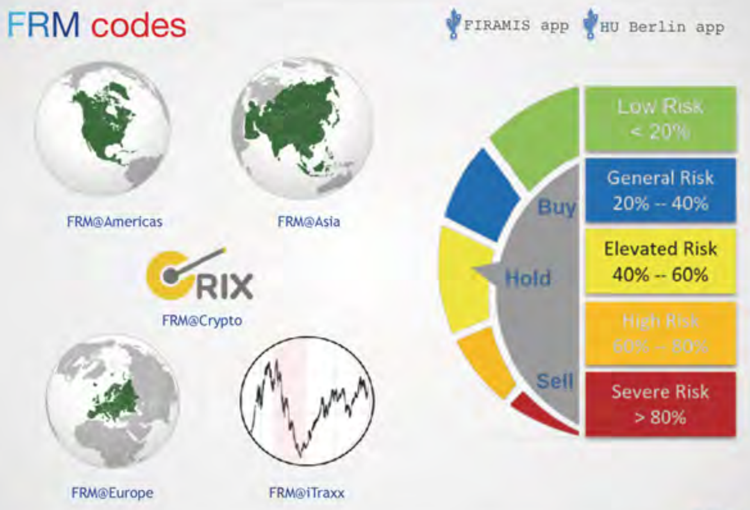
where: $df$ dimensionality of fitted model

- $\lambda$ as function of $j, t$
- Distribution of $\lambda$
- ID the TE drivers





FRM


FRM codes
FIRAMIS app
HU Berlin app
FRM@Americas
FRM@Asia
CRIX
FRM@Crypto
FRM@Europe
FRM@iTraxx
Buy
Hold
Sell
Low Risk
< 20%
General Risk
20% -- 40%
Elevated Risk
40% -- 60%
High Risk
60% -- 80%
Severe Risk
> 80%
FRM

## FRM@Crypto Data

- 15 largest cryptocurrencies
- 6 macro related variables
- Quantile level $\tau = 0.05, 0.10, 0.25, 0.50$
- Time window $s = 63, 21$
- Time frame: 2014–2020
- Macroeconomic risk factors:
  - US dollar index (average of USD vs main non-crypto currencies)
  - Yield level in USD (carry component for the drift)
  - VIX
  - CVIX (same as VIX, but on major fiat currencies)
  - S&P500

LQ Lasso Regression

## Methodology

- Obtain risk driver list of all historically active index members
- Download daily rates in same currency (USD)
- Sort market cap decreasingly (to select $J$ biggest risk drivers)
- Calculate returns
- On every trading day
  - Select $J$ biggest risk driver's returns over $s$ trading days
  - Attach returns of macroeconomic risk factors
  - Calculate $\lambda$ for all companies
  - Calculate average $\lambda$, etc.
  - Store active set

# FRM@Crypto Distribution

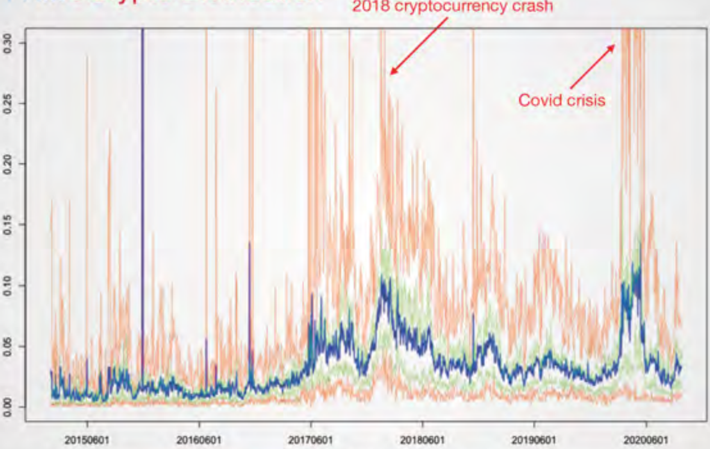



Figure: FRM@Crypto, Max and Min and 75 % and 25 % Quantiles for $\tau = 5\,\%$

# FRM@Crypto Distribution

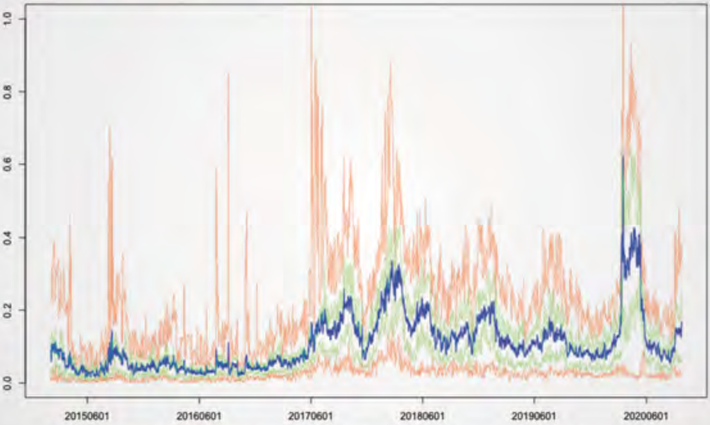



Figure: FRM@Crypto, Max and Min and 75 % and 25 % quantiles for $\tau = 10\,\%$

# FRM@Crypto Distribution

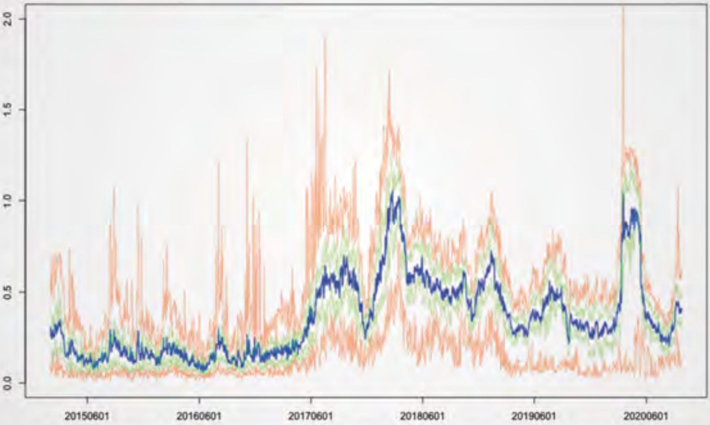



Figure: FRM@Crypto, Max and Min and 75 % and 25 % quantiles for $\tau = 25\,\%$

# FRM@Crypto Distribution

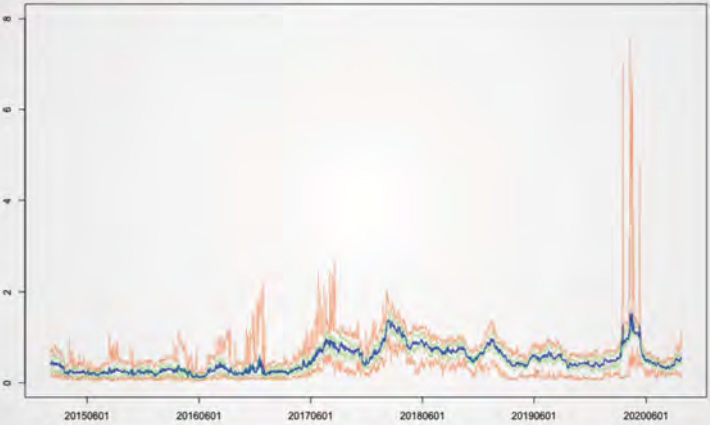
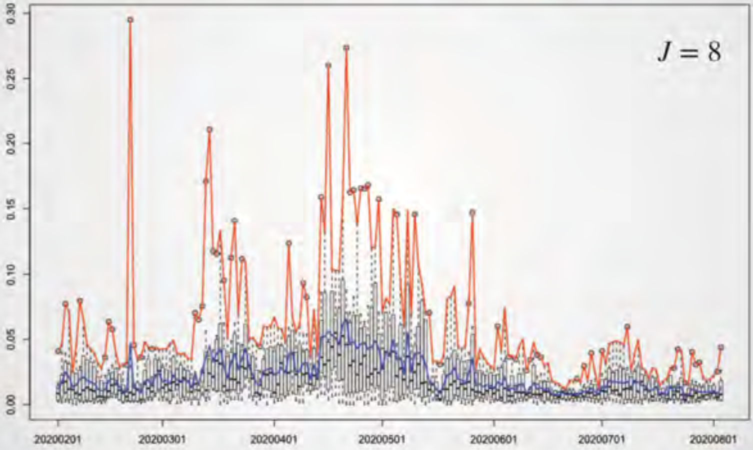
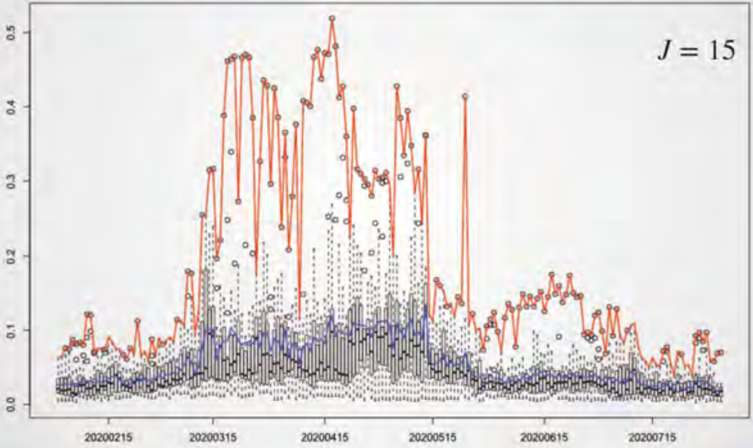



Figure: FRM@Crypto, Max and Min and 75 % and 25 % quantiles for $\tau = 50\,\%$

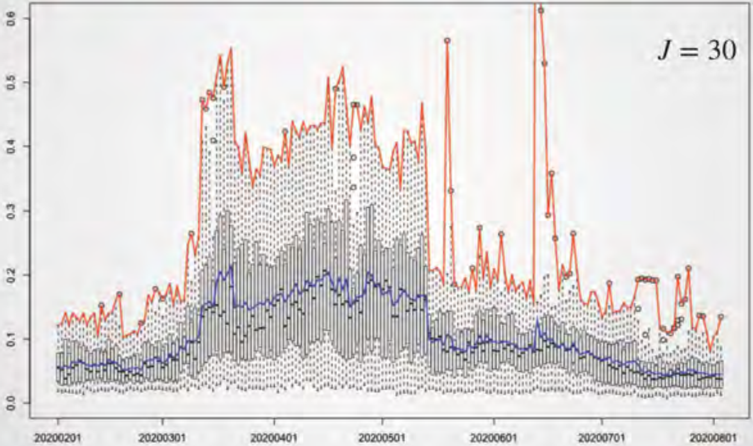
# BTC and ETH dominate the market - FRM reflects?



# BTC and ETH dominate the market - FRM reflects?



# BTC and ETH dominate the market - FRM reflects?



# Tail risk and window size sensitivity: FRM@Crypto Index

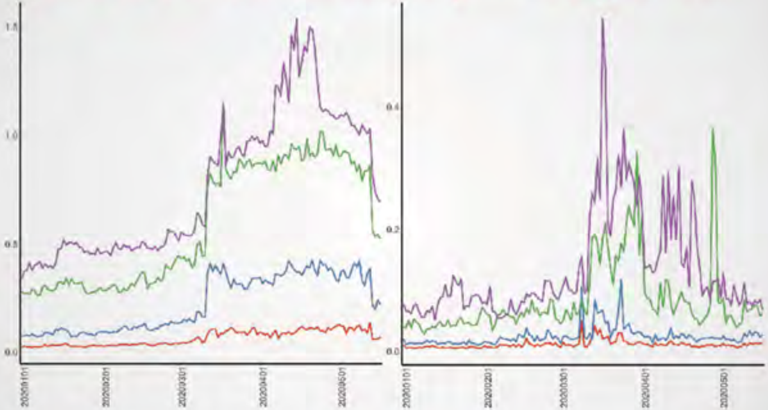


Figure: FRM@Crypto index for tail risk $\tau = 5\%, 10\%, 25\%, 50\%$ for $s = 63$ (left) and $s = 21$ (right). Data from 01 January 2020 to 17 May 2020.

# Tail risk and window size sensitivity: CoStress

| $\tau = 0.05$ | | $\tau = 0.10$ | | $\tau = 0.25$ | | $\tau = 0.50$ | |
|---|---|---|---|---|---|---|---|
| Crypto | Frequency | Crypto | Frequency | Crypto | Frequency | Crypto | Frequency |
| BTC | 112 | BTC | 95 | XRP | 93 | BTC | 116 |
| ETH | 76 | LTC | 83 | INNBCL | 90 | XRP | 97 |
| LTC | 61 | ETH | 63 | TAGZ5 | 80 | INNBCL | 87 |
| BSV | 57 | INNBCL | 57 | ETH | 73 | BSV | 65 |
| INNBCL | 57 | BCH | 44 | BTC | 70 | TAGZ5 | 65 |

| $\tau = 0.05$ | | $\tau = 0.10$ | | $\tau = 0.25$ | | $\tau = 0.50$ | |
|---|---|---|---|---|---|---|---|
| Crypto | Frequency | Crypto | Frequency | Crypto | Frequency | Crypto | Frequency |
| BCH | 74 | BCH | 79 | BCH | 81 | EOS | 85 |
| LINK | 74 | EOS | 63 | ADA | 79 | XMR | 84 |
| XMR | 70 | XMR | 63 | BNB | 73 | XLM | 81 |
| BNB | 64 | XTZ | 62 | XMR | 69 | BCH | 78 |
| EOS | 59 | ADA | 60 | XTZ | 65 | ADA | 72 |

Table: Crypto currencies with high (top table) and low (bottom table) CoStress with the number of days they appeared in top/bottom 5 for tail risk $\tau = 5\%, 10\%, 25\%, 50\%$.

Data from 1 January 2020 to 17 May 2020.

# FRM@Crypto Model Selection Methods

| | $\tau = 0.05$ | | | $\tau = 0.10$ | | | $\tau = 0.50$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\lambda_j$ | $SIC_j$ | $GACV_j$ | $\lambda_j$ | $SIC_j$ | $GACV_j$ | $\lambda_j$ | $SIC_j$ | $GACV_j$ |
| BTC | 0.039 | -6.633 | 0.001 | 0.266 | -5.928 | 0.002 | 0.837 | -4.387 | 0.011 |
| ETH | 0.018 | -2.495 | 0.074 | 0.030 | -1.974 | 0.130 | 0.702 | -1.644 | 0.193 |
| XRP | 0.034 | -6.871 | 0.001 | 0.368 | -5.683 | 0.003 | 0.828 | -4.702 | 0.008 |
| BCH | 0.071 | -6.742 | 0.001 | 0.307 | -5.947 | 0.002 | 0.763 | -4.794 | 0.007 |
| BSV | 0.058 | -6.686 | 0.001 | 0.292 | -5.773 | 0.003 | 0.504 | -5.011 | 0.006 |
| LTC | 0.052 | -6.396 | 0.001 | 0.097 | -5.747 | 0.003 | 0.729 | -4.516 | 0.010 |
| EOS | 0.030 | -4.516 | 0.009 | 0.064 | -4.128 | 0.015 | 0.402 | -3.609 | 0.025 |
| BNB | 0.088 | -7.206 | 0.001 | 0.039 | -6.633 | 0.001 | 0.929 | -4.459 | 0.011 |
| XTZ | 0.091 | -6.505 | 0.001 | 0.206 | -5.885 | 0.002 | 0.742 | -4.577 | 0.009 |
| LINK | 0.132 | -6.824 | 0.001 | 0.236 | -5.980 | 0.002 | 0.698 | -4.750 | 0.008 |
| ADA | 0.051 | -6.314 | 0.002 | 0.229 | -5.667 | 0.003 | 0.802 | -4.509 | 0.010 |
| XLM | 0.058 | -6.534 | 0.001 | 0.194 | -5.680 | 0.003 | 0.579 | -4.523 | 0.010 |
| XMR | 0.160 | -6.584 | 0.001 | 0.277 | -5.835 | 0.003 | 0.770 | -4.778 | 0.008 |
| TRX | 0.080 | -6.236 | 0.002 | 0.220 | -5.250 | 0.005 | 0.541 | -4.684 | 0.009 |
| HT | 0.059 | -6.513 | 0.001 | 0.226 | -5.716 | 0.003 | 0.526 | -4.679 | 0.008 |

Table: $\lambda_j$ and effective dimension of the $j$-fitted models, via solution path of the $L1$-norm QR algorithm and formula for $\tau = 5\%$, 10%, 50%. In all the settings $s = 63$, $J = 15$.
Data from 1 January 2020 to 17 May 2020.

# Flight into Cash: 2018 vs 2020 Crises

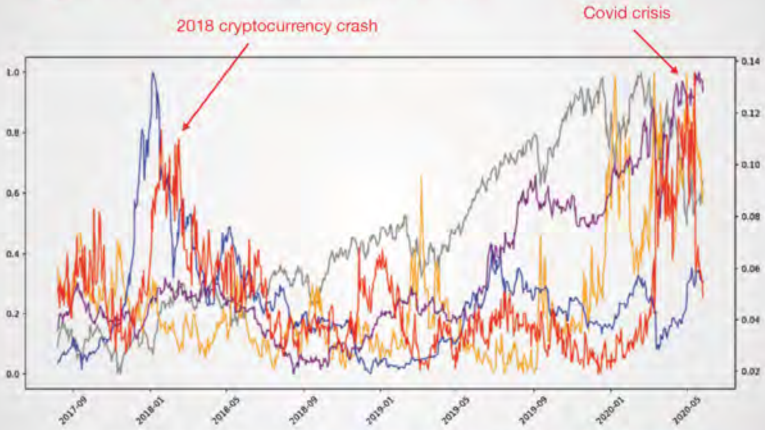



Figure: Normalised S&P500 Index, VIX Index, CRIX Index, Gold Price and FRM@Crypto

# FRM@Crypto Adjacency Matrix with Macro Variables

- $\tau = 0.05$, 12 February 2018

| | BTC | ETH | XRP | BCH | ADA | LTC | NEO | XLM | EOS | MIOTA | XEM | DASH | XMR | LSK | TRX | 1Y | CVIX | DXY | SPX | VIX | VCRIX |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BTC | | | 0.13 | | 0.04 | 0.10 | 0.00 | | 0.04 | 0.07 | -0.12 | | 0.13 | | 0.00 | | | | | -0.11 | 0.17 |
| ETH | | | 0.03 | 0.07 | | 0.24 | 0.10 | | | 0.01 | | 0.04 | | 0.13 | 0.02 | | | | | -0.14 | |
| XRP | | | | 0.33 | -0.03 | | -0.03 | 0.35 | 0.07 | | 0.17 | | | -0.13 | | | | | | 0.04 | 0.14 |
| BCH | | 0.18 | -0.03 | | | | 0.08 | | | -0.05 | 0.00 | 0.45 | 0.32 | | 0.01 | | | | | 0.08 | |
| ADA | | | | | | | | | | | | | | | | | | | | | |
| LTC | 0.26 | 0.23 | | | | | | | 0.02 | 0.16 | 0.00 | | -0.01 | | | | | | | | |
| NEO | | | 0.07 | 0.24 | 0.00 | 0.18 | 0.23 | 0.02 | | 0.15 | 0.01 | | | | 0.02 | | | | | | |
| XLM | | | | | | | | | | | | | | | | | | | | | |
| EOS | | | | | | | | | | | | | | | | | | | | | |
| MIOTA | | | | | | | | | | | | | | | | | | | | | |
| XEM | | 0.12 | 0.19 | 0.04 | | 0.06 | 0.10 | 0.19 | | | 0.13 | | | | 0.06 | | | | | | |
| DASH | | | 0.10 | 0.12 | 0.40 | | | | | 0.04 | 0.07 | | 0.25 | | -0.14 | | | | | | |
| XMR | | | 0.01 | 0.23 | 0.10 | | 0.18 | | | 0.08 | | | | 0.05 | 0.02 | | | | | | |
| LSK | 1.12 | | 0.06 | 0.20 | | | -0.52 | -0.03 | | | | 0.11 | 0.16 | | | | | | | | 0.26 |
| TRX | | | | | | | | | | | | | | | | | | | | | |

Few traditional macro variables explain crypto currency tail behaviour



FRM

# Visualising the Active Set: FRM@Crypto the Movie

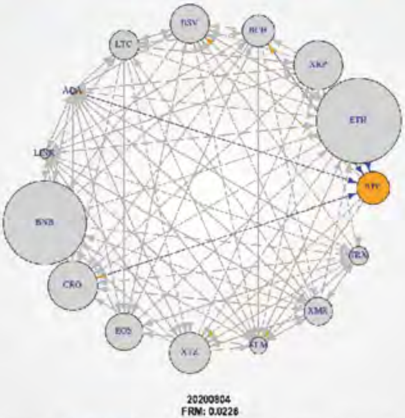
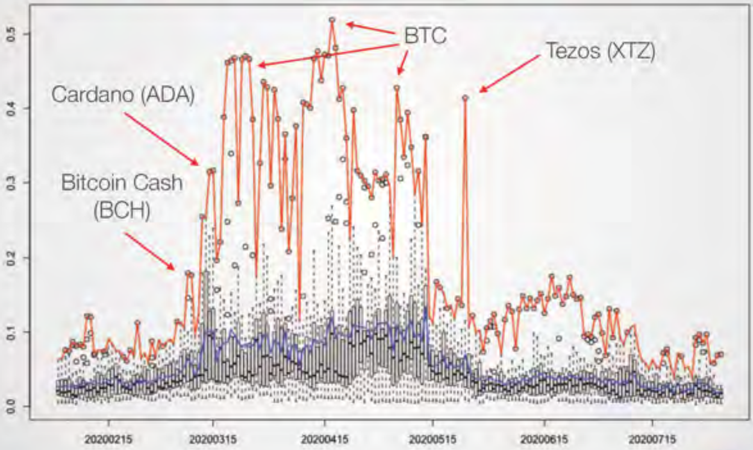
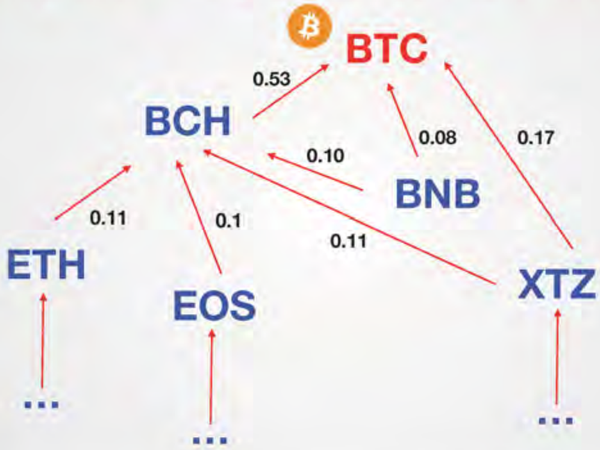



Figure: Network analysis for FRM@Crypto from 4 August 2020 to 24 September 2020. Size of the node corresponds to $\lambda$

# FRM@Crypto Distribution under Covid Crisis



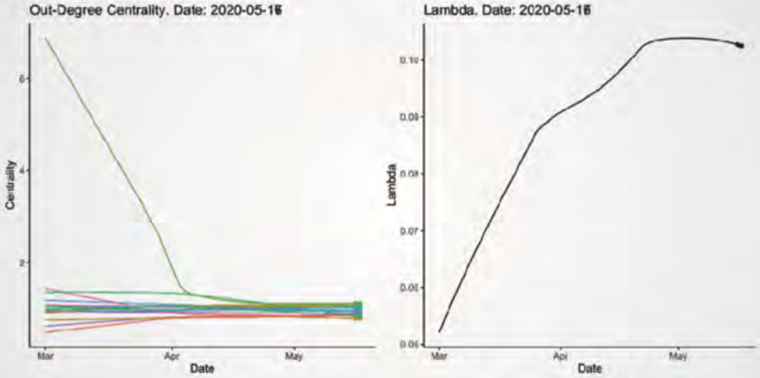
## 29 April 2020 — Marginal Return Contribution to BTC



## Types of Centrality of a Node

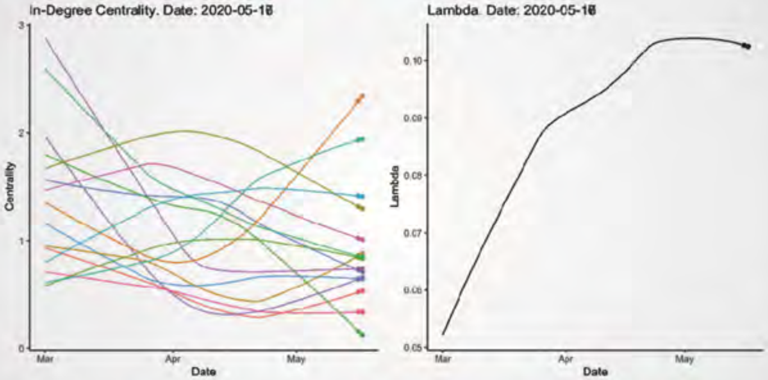
- Degree centrality
  - *In-degree* — how many other coins affect the node
  - *Out-degree* — how many other coins the node affects
- *Closeness* — shortest path between the node and all other nodes
- *Betweenness* — the number of times a node acts as a bridge along the shortest path between two other nodes
- *Eigenvector* — takes into account that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes

# FRM@Crypto Out-Degree Centrality



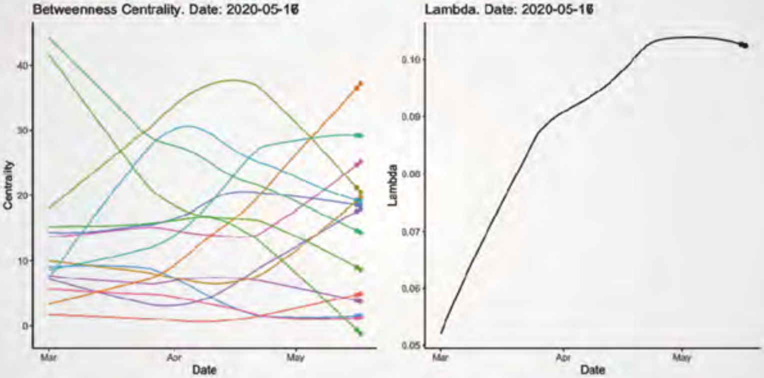

Left-hand side panel: # of outbounds links of BTC, ETH, XRP, BCH, BSV, LTC, EOS, BNB, XTZ, LIN, ADA, XLM, XMR, TRX, HT. Right-hand side panel: FRM index over time. Data from 01 March 2020 to 17 May 2020

# FRM@Crypto In-Degree Centrality



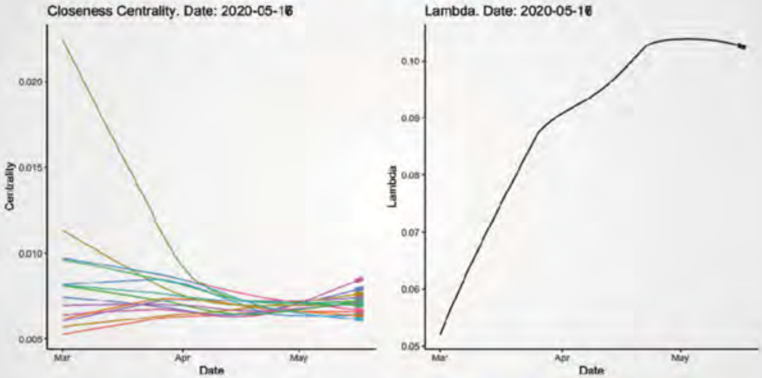

Left-hand side panel: # of inbound links of BTC, ETH, XRP, BCH, BSV, LTC, EOS, BNB, XTZ, LIN, ADA, XLM, XMR, TRX, HT. Right-hand side panel: FRM index over time.
Data from 01 March 2020 to 17 May 2020

# FRM@Crypto Betweenness Centrality



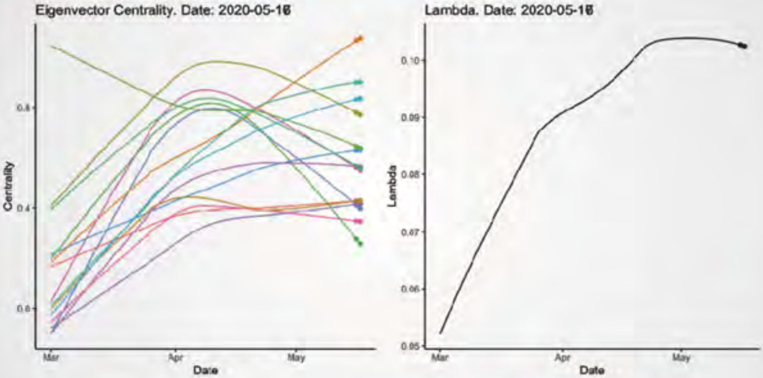

Left-hand side panel: „bridge" behaviour measure for BTC, ETH, XRP, BCH, BSV, LTC, EOS, BNB, XTZ, LIN, ADA, XLM, XMR, TRX, HT. Right-hand side panel: FRM index over time. Data from 01 March 2020 to 17 May 2020

# FRM@Crypto Closeness Centrality




Left-hand side panel: fastness in influencing of BTC, ETH, XRP, BCH, BSV, LTC, EOS, BNB, XTZ, LIN, ADA, XLM, XMR, TRX, HT. Right-hand side panel: FRM index over time. Data from 01 March 2020 to 17 May 2020

# FRM@Crypto Eigenvector Centrality




Left-hand side panel: normalised eigenvector centrality of BTC, ETH, XRP, BCH, BSV, LTC, EOS, BNB, XTZ, LIN, ADA, XLM, XMR, TRX, HT. Right-hand side panel: FRM index over time. Data from 01 March 2020 to 17 May 2020

## From Nodes to Network Centralisation

Extend the notion of *point centrality* on the entire network.

1. Average of all nodes > spirit of FRM

$$C = \sum_{i=1}^{M} C(p_i)$$

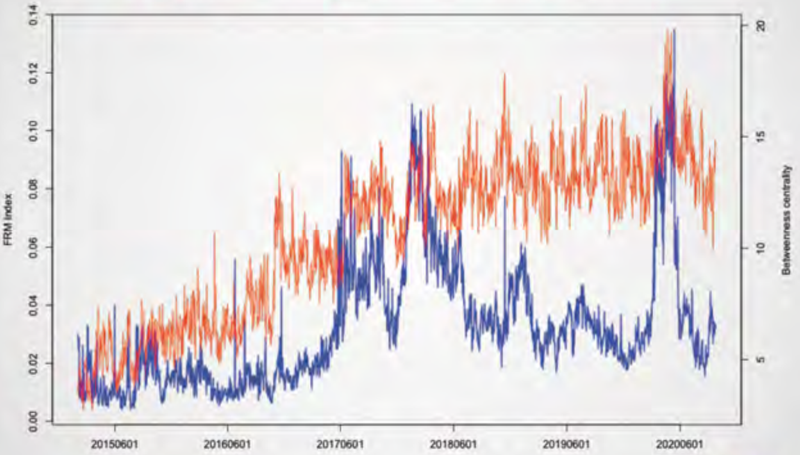
2. Freeman centralisation

$$C = \frac{\sum_{i=1}^{M} [C(p_*) - C(p_i)]}{\max \sum_{i=1}^{M} [C(p_*) - C(p_i)]} \underset{\text{star}}{=} \frac{\sum_{i=1}^{M} [C(p_*) - C(p_i)]}{M^2 - 3M + 2}$$

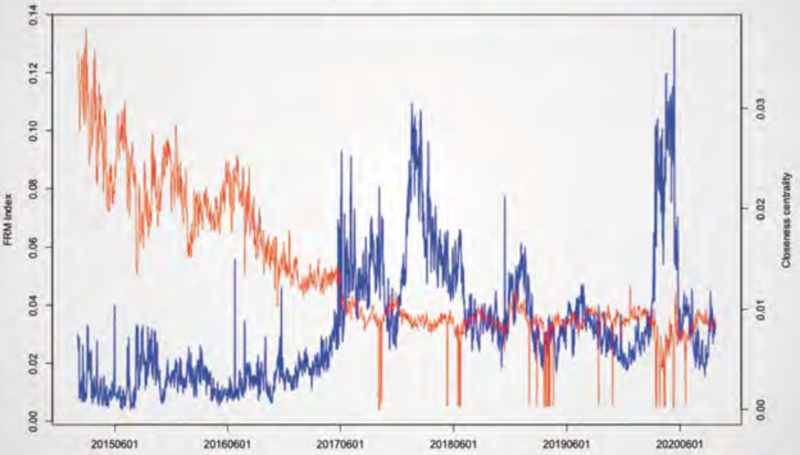
$p_*$ is most central node, max is over all graphs with $M$ nodes.

# FRM@Crypto vs Average Degree Centrality

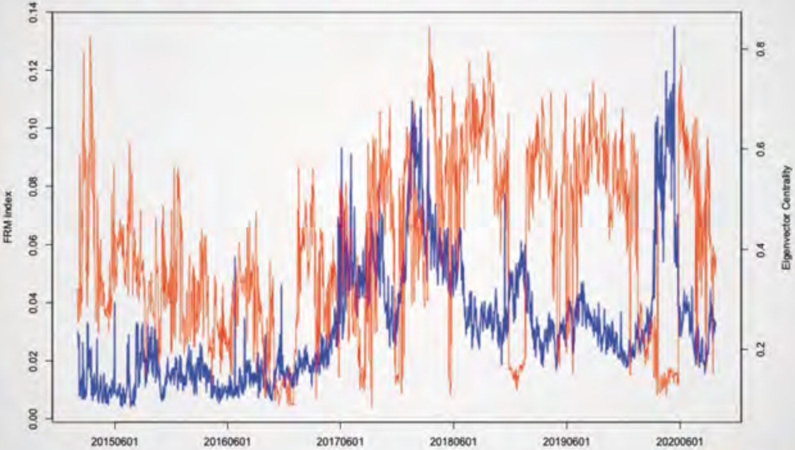
# FRM@Crypto vs Average Betweenness Centrality



# FRM@Crypto vs Average Closeness Centrality



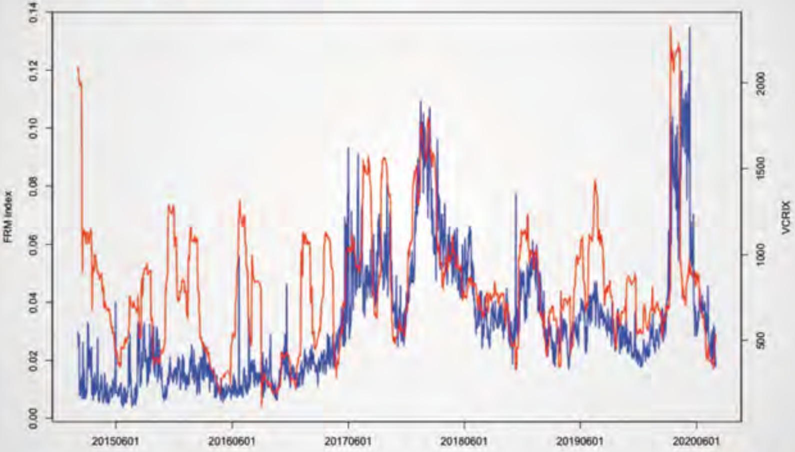
## FRM@Crypto vs Average Eigenvector Centrality



## Backtesting

- Assess model validity based on the usefulness of its predictions and not on the sophistication of the assumptions
- How well the risk measured by individual lambdas or their average reflects the short-time riskiness of cryptos
  - Riskiness benchmark: rolling historical volatility
  - Estimation window: 63 days

# FRM@Crypto Index and VCRIX



# Graphical backtest, $\tau = 5\,\%$

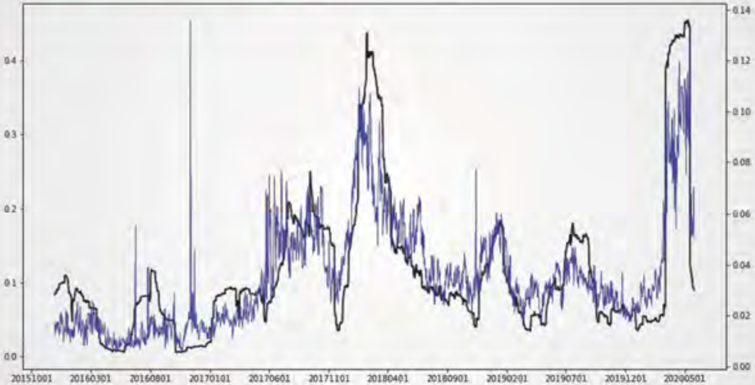



Figure: FRM@Crypto for $\tau = 5\,\%$ and CRIX rolling variance 20150404–20200525

# Graphical backtest, $\tau = 10\,\%$

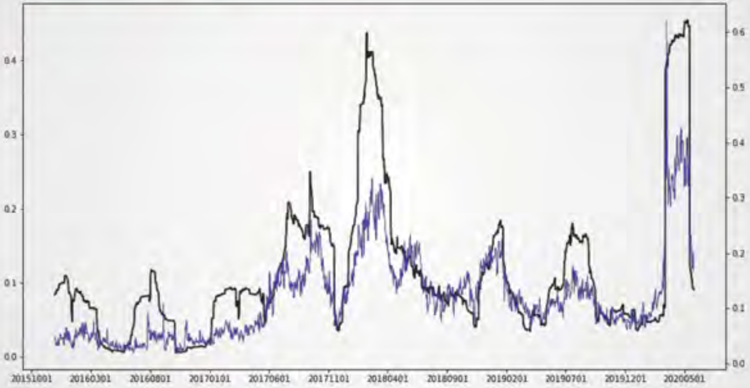



Figure: FRM@Crypto for $\tau = 10\,\%$ and CRIX rolling variance 20150404–20200525

# Graphical backtest, $\tau = 5\,\%$

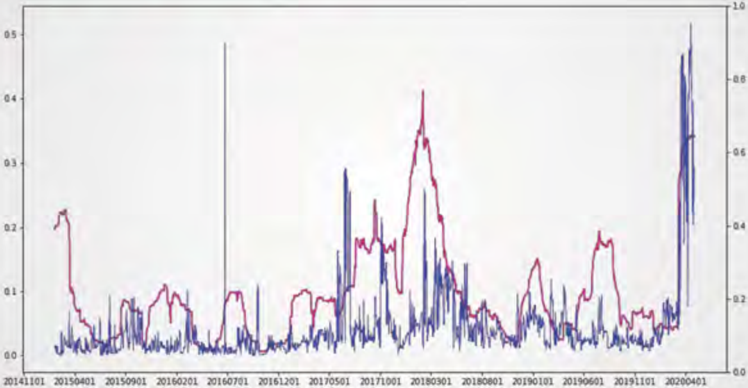



Figure: BTC lambda for $\tau = 5\,\%$ and BTC rolling variance 20150201-20200428

## Graphical backtest, $\tau = 10\,\%$

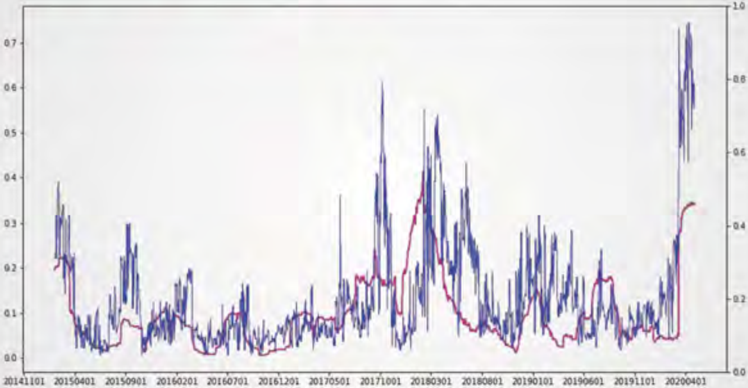



Figure: BTC lambda for $\tau = 10\,\%$ and BTC rolling variance 20150201–20200428

## Graphical backtest, $\tau = 5\,\%$

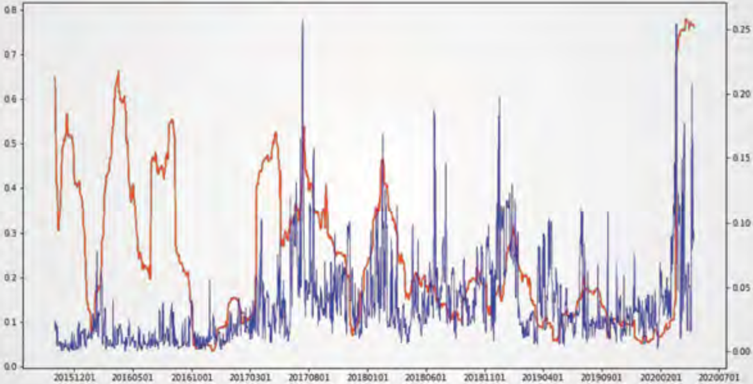



Figure: ETH lambdas for $\tau = 5\,\%$ and ETH rolling variance 20151011–20200428

# Graphical backtest, $\tau = 10\,\%$

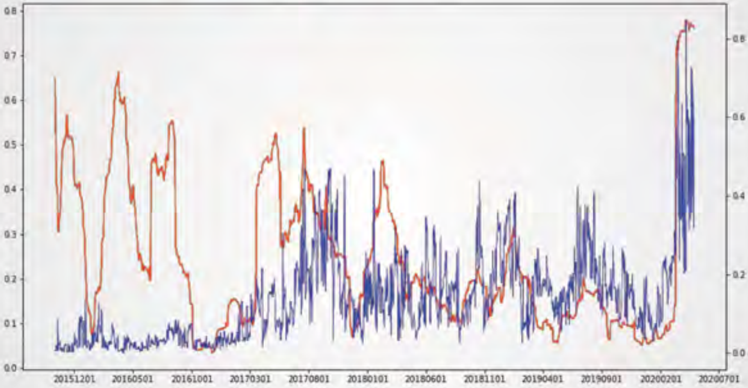



Figure: ETH lambdas for $\tau = 10\,\%$ and ETH rolling variance 20151011–20200428

## Crypto Returns and the Pricing Kernel

According to the basic pricing equation

$$\mathrm{E}_t^{\mathbb{P}}[m_{t+1} R_{i,t+1}] = 1 \tag{4}$$

$m_t$ *marginal rate of substitution*, $R_{i,t}$ return of $i$-th crypto.

Considering log returns $r_{i,t} = \log(R_{i,t}) \approx R_{i,t} - 1$

$$\mathrm{E}_t^{\mathbb{P}}[(1 + r_{i,t+1})\, m_{t+1}] \approx 1 \tag{5}$$

Substituting in (4) risk-free rate $R_{i,t} = R^f$

$$\mathrm{E}_t^{\mathbb{P}}[m_{t+1}] = 1 \tag{6}$$

## Link to Sharpe Ratio

Combining (5) and (6)

$$E_t^{\mathbb{P}}[m_{t+1} r_{i,t+1}] \approx 0 \tag{7}$$

Hence, the „Sharpe" ratio of $r_{i,t}$ is bounded by $\sigma(m_t)$

$$(7) \;<>\; \underbrace{E_t^{\mathbb{P}}[m_{t+1}]}_{=1} E_t^{\mathbb{P}}[r_{i,t+1}] + \mathrm{Cov}_t^{\mathbb{P}}[m_{t+1} r_{i,t+1}] \approx 0$$

$$<>\; E_t^{\mathbb{P}}[r_{i,t+1}] \approx -\underbrace{\mathrm{Corr}_t^{\mathbb{P}}[m_{t+1} r_{i,t+1}]}_{\in[-1,1]} \sigma(m_{t+1})\,\sigma(r_{i,t+1})$$

$$> \; \left| E_t^{\mathbb{P}}[r_{i,t+1}] \right| \leq \sigma(m_{t+1})\,\sigma(r_{i,t+1}) \tag{8}$$

## Role of Lambda as Penalisation Parameter

An analogous inequality to (8) holds for the empirical distribution

$$\left| \widehat{\mathrm{E}}_t^{\mathbb{P}}[r_{i,t+1}] \right| \leq \sigma(\widehat{m}_{t+1})\, \widehat{\sigma}(r_{i,t+1})$$

$$\succ \left| \mathrm{E}_t^{\mathbb{P}}[r_{i,t+1}] \right| - \left| \widehat{\mathrm{E}}_t^{\mathbb{P}}[r_{i,t+1}] \right| \leq \sigma(m_{t+1})\, \sigma(r_{i,t+1}) - \sigma(\widehat{m}_{t+1})\, \widehat{\sigma}(r_{i,t+1}) \qquad (9)$$

Due to persistency of volatility of returns $\widehat{\sigma}(r_{i,t+1}) \approx \sigma(r_{i,t+1})$

$\lambda_{i,t}$ chosen with CV tries to minimise the LHS $\succ$

$$\lambda_{i,t} \propto \sigma(\widehat{m}_{t+1}) - \sigma(m_{t+1})$$

## FRM in FinTech, Cryptos, ...


Vol 1. 2019 on Crypto Currencies


Michael Althof


Vanessa Guarino


Wolfgang K Härdle


Rui REN


Anna Shchekina


Alla Petukhina


Ang LI


Souhir Ben Amor


Alex Truesdale


Ilyas Agakishiev

## References


- Adrian J, Brunnermeier M (2016) CoVaR, American Economic Review, 106 (7): 1705-41. DOI: 10.1257/aer.20120555
- Buraschi A, Corielle F (2005). Risk management of time-inconsistency: Model updating and recalibration of no-arbitrage models. J Banking and Finance 29: 2883-907
- Chao SK, Härdle WK, Wang W (2015) Quantile Regression in Risk Calibration. Handbook for Financial Econometrics and Statistics, Cheng-Few Lee, ed., Springer Verlag, DOI: 10.1007/978-1-4614-7750-1_54.
- Härdle WK, Wang W, Zbonakova L (2018) Time Varying Lasso, in Applied Quantitative Finance 3rd ed. (Chen, Härdle, Overbeck eds.) Springer Verlag, ISBN 978-3-662-54486-0
- Keilbar G (2018) Modeling systemic risk using Neural Network Quantile Regression, MSc thesis
- Li Y, Zhu JL (2008) L1 Norm Quantile Regression, J Comp Graphical Statistics 17(1): 1-23
- Osborne MR, Presnell B, Turlach BA (200) J Comp Graphical Statistics Vol. 9, 319-337
- Yuan, M. (2006), GACV for Quantile Smoothing Splines, Computational Statistics & Data Analysis, 50: 813{829

# FRM financialriskmeter for Cryptos

Michael Althof

Vanessa Guarino

Rui Ren

Anna Shchekina

Humboldt-Universität zu Berlin
lvb.wiwi.hu-berlin.de
Charles University, WISE XMU, NCTU 玉山学者

## Expectile as Quantile

$e_\tau(Y)$ is the $\tau$-quantile of the cdf $T$, where

$$T(y) = \frac{G(y) - x\,F(y)}{2\{G(y) - y\,F(y)\} + y - \mu_Y}$$

and

$$G(y) = \int_{-\infty}^{y} u\,dF(u)$$

▶ Back to Expectiles

# Cryptocurrencies List (as per 24 May 2020)

| Symbol | Name | Last Price (USD) | Market Cap (USD) | 24H Volumes (USD) |
|---|---|---|---|---|
| BTC | Bitcoin | 8946.62 | 164481372045 | 27576284769 |
| ETH | Ethereum | 203.41 | 22618375461 | 9311268064 |
| XRP | XRP | 0.19 | 8625857668 | 1236573262 |
| BCH | Bitcoin Cash | 226.73 | 4175489941 | 2639464553 |
| BSV | Bitcoin SV | 189.55 | 3492449683 | 939543182 |
| LTC | Litecoin | 42.79 | 2777753749 | 2307602277 |
| EOS | EOS | 203.46 | 22568743176 | 9923363991 |
| BNB | Binance Coin | 16.17 | 2393754841 | 258305237 |
| XTZ | Tezos | 2.70 | 1923243499 | 82421482 |
| LINK | ChainLink | 3.87 | 1469368639 | 358145283 |
| ADA | Cardano | 0.053 | 1656068633 | 100244607 |
| XLM | Stellar | 0.066 | 1333292859 | 323203952 |
| XMR | Monero | 62.03 | 1089971286 | 91193644 |
| TRX | TRON | 0.015 | 970220373 | 1372904826 |
| HT | Huobi Token | 8947.42 | 164496303531 | 27970959275 |

Source: www.coingecko.com

▶ FRM equations