

RESEARCH ARTICLE

Transfer entropy as a variable selection methodology of cryptocurrencies in the framework of a high dimensional predictive model

Andrés García-Medina^{1,2*}, Graciela González Farías³

1 Consejo Nacional de Ciencia y Tecnología, Av. Insurgentes Sur 1582, Col. Crédito Constructor 03940, Ciudad de México, México, **2** Unidad Monterrey, Centro de Investigación en Matemáticas, A.C. Av. Alianza Centro 502, PIIT 66628, Apodaca, Nuevo Leon, Mexico, **3** Probability and Statistics, Centro de Investigación en Matemáticas, A.C. Jalisco S/N, Col. Valenciana 36240, Guanajuato, Mexico

* andres.garcia@cimat.mx



OPEN ACCESS

Citation: García-Medina A, González Farías G (2020) Transfer entropy as a variable selection methodology of cryptocurrencies in the framework of a high dimensional predictive model. PLoS ONE 15(1): e0227269. <https://doi.org/10.1371/journal.pone.0227269>

Editor: J E. Trinidad Segovia, University of Almeria, SPAIN

Received: April 28, 2019

Accepted: December 15, 2019

Published: January 2, 2020

Copyright: © 2020 García-Medina, González Farías. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by FOSEC SEP- INVESTIGACIÓN BÁSICA through Consejo Nacional de Ciencia y Tecnología, México, 252996 (<https://www.conacyt.gob.mx/index.php/sni/convocatorias-conacyt/convocatorias-fondos-sectorialesconstituidos/convocatoria-sep-conacyt/investigacion-basica-sep>, GG and AG); and FOSEC

Abstract

We determine the number of statistically significant factors in a high dimensional predictive model of cryptocurrencies using a random matrix test. The applied predictive model is of the reduced rank regression (RRR) type; in particular, we choose a flavor that can be regarded as canonical correlation analysis (CCA). A variable selection of hourly cryptocurrencies is performed using the Symbolic estimation of Transfer Entropy (STE) measure from information theory. In simulated studies, STE shows better performance compared to the Granger causality approach when considering a nonlinear system and a linear system with many drivers. In the application to cryptocurrencies, the directed graph associated to the variable selection shows a robust pattern of predictor and response clusters, where the community detection was contrasted with the modularity approach. Also, the centralities of the network discriminate between the two main types of cryptocurrencies, i.e., coins and tokens. On the factor determination of the predictive model, the result supports retaining more factors contrary to the usual visual inspection, with the additional advantage that the subjective element is avoided. In particular, it is observed that the dynamic behavior of the number of factors is moderately anticorrelated with the dynamics of the constructed composite index of predictor and response cryptocurrencies. This finding opens up new insights for anticipating possible declines in cryptocurrency prices on exchanges. Furthermore, our study suggests the existence of specific-predictor and specific-response factors, where only a small number of currencies are predominant.

Introduction

In econometrics, it is of fundamental interest to determine the proper number of components in a multivariate model because this allows for the attribution of explanatory meaning to each factor based on economic theory. Traditionally, a visual inspection approach has been a

SEP-INVESTIGACIÓN BÁSICA through Consejo Nacional de Ciencia y Tecnología, México, A1-S-43514 (<https://www.conacyt.gob.mx/index.php/sni/convocatorias-conacyt/convocatorias-fondos-sectorialesconstituidos/convocatoria-sep-conacyt/investigacion-basica-sep>, AG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

standard methodology in factor analysis and principal component analysis (PCA) since the seminal publication [1]. There, a technique known as the scree test, whereby the eigenvalues associated with the covariance matrix are ordered from largest to smallest and plotted as a downward curve, was proposed. According to the test, we must look for an *elbow* in the curve and retain the factors associated with the eigenvalues to the left of that point. Of course, this methodology has the disadvantage of being highly subjective, and different researchers can choose different numbers of factors. Another usual approach originates from purely nonparametric statistics and relies on the cross-validation technique [2]. Methods of this type are computationally demanding and become impractical if the number of variables increases at the same rate as the number of observations in the model.

A more recent approach using a combination of parametric and nonparametric assumptions has been proposed in [3]. The researchers include as significant factors quantities for which the mean eigenvalue distribution under a bootstrapping resampling is larger than one. The latter criterion is attributed to Kaiser [4] but was proposed analytically by Guttman [5]. This method is also biased if the number of variables increases with that of observations. In this case, it is better to use a modified version of the Kaiser approach [6] to perform analyses in this regime known as high-dimensional statistics. In high-dimensional factor analysis, there are relevant proposals for determining the number of factors that originate in random matrix theory [7, 8]. Even in the related estimation method based on ratio tests of eigenvalues [9], some conclusions are consequences of the results for random matrices. In the standard use of random matrices, the Tracy-Widom distribution [10] plays a crucial role in testing the sample eigenvalue distribution. Additionally, more sophisticated approaches use the joint Tracy-Widom distribution [8] to determine the number of factors, and further elegant versions use free matrices and noncommutative probability [11, 12]. Nevertheless, all the novel approaches in the high-dimensional regime are mainly focused on PCA and factor analysis, while other multivariate models such as the canonical correlation analysis (CCA) have drawn less attention.

In this study, we are interested in a model of the latter type, where CCA is presented as a particular case of the reduced-rank regression model. In this case, Johnstone [13] proved that the Tracy-Widom distribution could be used to determine the number of significant factors after appropriate transformations of the greatest root distribution involved in the Union Intersection Test (UIT). Based on this technique, the idea is to test the null hypothesis $H_0: \Sigma_{xy} = I$, where Σ_{xy} represents the covariance matrix between Gaussian datasets \mathbf{X} and \mathbf{Y} . In some sense, not imposing a structure simplifies the analysis. However, it is important to keep in mind that we will follow a parametric approach, and some restrictions are imposed accordingly. The relevant assumption here is that the data must follow a normal distribution, which usually is untrue for financial time series in the high-frequency domain [14]. Considering this problem, Burda et al. [15, 16] have derived a heavy-tail limit distributions of eigenvalues based on the framework of random matrices. The researchers determined that in the limit case of Levy processes, the distribution of eigenvalues of the sample correlation matrix does not have a bounded support. Hence, no limit distribution of the largest eigenvalues exists as an equivalent to the Tracy-Widom distribution. Therefore, the results obtained in this study can be thought as an upper bound for the maximum number of significant factors, and the true number can be less if the relevant time series is heavy-tailed. But as an intermediate case, if the density of the matrix entries behaves like $|x|^{-\mu}$, then there exist physical arguments to support a phase transition from Tracy-Widom to Poisson at $\mu = 4$ [17]. Then, this enable to use the Tracy-Widom statistics to financial time series with moderate heavy-tail behavior.

Thus, our intention is to implement the Tracy-Widom test to determine the number of significant factors in a predictor-response set of cryptocurrencies modeled by CCA. These new

financial instruments are based on blockchain [18] technology, where a coin is defined as a chain of digital signatures. Each owner transfers the coin to the next owner by digitally signing a hash of the previous transaction and the public key of the next owner and adding these to the end of the coin. The easy access to this new financial instrument through more than 17000 exchanges with low transaction fees, more than 2000 virtual currencies worldwide and a traded volume of nearly 60 billion dollars has made it a very attractive investment instrument for the general population [19].

Interestingly, several studies have shown evidence that such instruments are inefficient in the meaning of the efficient market hypothesis [20]. Consequently, predictive strategies can be applied to earn profits from trading on the relevant virtual exchange platforms. To mention some examples, in [21], Bitcoin was studied over the historical period of 2010-2016 with a battery of robust tests, and the study concluded that Bitcoin was transitioning from an inefficient market to an efficient one. Another study [22] has exploited this inefficiency based on a machine learning framework and searched for abnormal profits using a representative set of cryptocurrencies traded on various exchanges during 2015-2018. Additionally, the above study presented several insights for the prediction of the short-term evolution of the cryptocurrency market. Another study [23] in this direction observed persistence in four major cryptocurrencies, i.e., a positive correlation between a cryptocurrency's past and future values, and similarly concluded that unusual profits could be earned by trading in these markets. Hence, given the above studies, it does not seem unreasonable to study the statistical determination of the number of factors in a multivariate predictive model, e.g., using CCA.

Similarly, attempts have been made to characterize the collective behavior of cryptocurrencies; one example is [24]. There, it was shown that a large dataset of cryptocurrencies at the daily frequency deviated from the universal results of Marchenko-Pastur [25]. Additionally, the study stated that the spanning tree structure was stable over time. Furthermore, the power-law behavior of Bitcoin was analyzed in [26] over a long period of time and at various frequencies from one minute to one day. The researchers concluded that Bitcoin exhibited heavy tails in the range of $2 < \alpha < 2.5$ across multiple coin exchanges. Their findings supported the use of standard financial methods because of the finite variance implications of results. Nonetheless, we are interested in remaining within the domain of the parametric approach of random matrices since the computational cost is minimal and only a few assumptions are required.

On the other hand, a crucial step in any predictive model is the variable selection procedure used to define the predictor and response variables. In most cases, this is done based on the accumulated research experience or according to the consensus of experts in the field. However, cryptocurrencies are a new financial instrument for which there is a limited amount of previous experience in making this selection. Thereby, this study proposes using the transfer entropy measure to solve the variable selection problem. Transfer entropy is a dynamic and nonsymmetric measure that was initially developed by Schreiber [27] and is based on the concept of Shannon entropy [28]. This measure was designed to determine the directionality of transfer information between two processes by detecting the asymmetry in their interactions [29]. Transfer entropy has been used to solve numerous problems. It has been useful in the study of the neuronal cortex of the brain [30], statistical physics [31], and dynamic systems [32], and was given a thermodynamic interpretation in [33]. In applications to econometrics, transfer entropy can be regarded as a nonlinear generalization of the Granger causality test [34]. In this field, effective transfer entropy [35] has been proposed for dealing with finite sample effects. Nevertheless, it does not have an empirical limit distribution that can be used as a comparison. Instead, effective transfer entropy is based in resampling and the use of surrogate data. The study of Sandoval [36] uses this approach to study the contagion of institutions in times of crisis. The cited study identifies the companies most vulnerable to contagion and

dependent on failing economies. A more recent study in this direction is [37], where transfer entropy is estimated by discretizing the return time series into positive and negative unit values. The researchers used stock market and real estate data to construct an indicator used to measure systematic risk to predict future market volatility.

In the above applications, the estimation method is based on binning. Our intention is to use the symbolic approach, where the time series can be thought of as embedded in a dynamic system. This avoids fine-tuning of parameters, which usually limits the use of transfer entropy to field applications. Symbolic transfer entropy is a robust and computationally fast method of quantifying the dominant direction of information flow between time series [38]. Furthermore, in [39], the multivariate version of symbolic transfer entropy has been tested, and it has been shown that it can be applicable to nonstationary time series in mean and variance and is even unaffected by the existence of outliers and vector autoregressive filtering. Moreover, another advantage of using the symbolic approach is that under some circumstances, there exists a null hypothesis regarding the distribution that can be used to measure the absence of a direct flow of information. This makes the results more robust and simpler to compute.

To end this section, let us point out that, in this study, our concerns are focused on the proper determination of the number of components and variable selection beyond the predictive precision of the CCA model. Thus, the aim of this study is twofold. First, it is to provide tools related to variable selection and the number of statistically significant factors in a multiresponse forecast obtained with CCA that describes an apparently unrelated mathematical apparatus. Second, it is to explore the methodology's application to the new cryptocurrency instruments. However, the proposed approach is general, and this framework can be applied to the analysis of any dataset of interest.

In the next section, the preprocessing of the dataset of cryptocurrencies is described. Next, in the section on variable selection, the transfer entropy measure is used to discriminate between the set of predictor and response variables, and the results for cryptocurrency-related variables are presented. The regression framework section introduces the general regression model that serves as the setting for the studied model. Then, in the section on the number of factors, random matrix theory is used to select the appropriate number of factors in the presented multiresponse regression model considered at a high-dimensional setting. The mathematical relation of results in high-dimensional statistics with the reduced-rank selection problem for the particular case of CCA is also described there. Next, the consequences of the methodology are explored by considering the set of predictor-response cryptocurrencies. Finally, in the concluding section, the main findings are summarized, and future research directions are proposed.

Data

A sample of $p = 100$ cryptocurrencies is obtained using the API of CoinMarketCap [19] for the period from May 23 to November 27, 2018, at an hourly frequency, resulting in a total of $n = 4533$ observations (see [S1 File](#) and [S1 Table](#)).

We transform exchange rates in dollars $P_k(t)$ to returns $R_k(t)$ for every cryptocurrency ($k = 1, \dots, p$) and moment of time ($t = 1, \dots, n$)

$$R_k(t) = \frac{Z_k(t + \Delta t) - Z_k(t)}{Z_k(t)}, \quad (1)$$

and analyze the standardized returns $r_k = (R_k - \mu_k)/\sigma_k$ ($k = 1, \dots, p$), where σ_k is the standard deviation of R_k , and μ_k denotes the average over time for the studied period. Due to this

transformation, the augmented Dickey-Fuller test [40] confirms that the relevant time series are stationary, with a *p-value* of less than 0.01 for all r_k ($k = 1, \dots, p$) considered.

Variable selection

One of the first problems encountered when trying to establish a predictive model is variable selection. In the econometric approach, the economic theory usually dictates which variables must be treated as predictors and as responses. However, cryptocurrencies are a new financial instrument with not many economic models available. Hence, we follow an information-theoretic approach to solve the variable selection problem.

In 2000, T. Schreiber introduced a quantity called transfer entropy (TE) in the context of information theory, with the purpose of measuring the information flow from one process to another in a nonsymmetrical way. Let $x_i = x(i)$ and $y_i = y(i)$, $i = 1, \dots, N$; denote sequences of observations of systems X and Y . TE is defined as [27]

$$T_{Y \rightarrow X}(k, l) = \sum_{ij} p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log \frac{p(x_{t+1}|x_t^{(k)}, y_t^{(l)})}{p(x_{t+1}|x_t^{(k)})}, \tag{2}$$

The idea behind TE is to incorporate time dependence by relating previous samples x_i and y_i to predict the next value x_{i+1} and quantify the deviation from the generalized Markov property, $p(x_{i+1}|x_i, y_i) = p(x_{i+1}|x_i)$, where p denotes the transition probability density. If there is no deviation from the generalized Markov property, Y has no influence on X . TE, formulated as the Kullback-Leibler entropy [41] between $p(x_{i+1}|x_i, y_i)$ and $p(x_{i+1}|x_i)$, quantifies the incorrectness of this assumption and is explicitly nonsymmetric with respect to the exchange of x_i and y_i .

An interesting property of TE is that under some conditions it can be regarded as a nonlinear generalization of Granger causality. In econometrics, Granger causality plays an important role in parameter estimation of a vector autoregressive (VAR) model. Consider the jointly stationary stochastic processes X_t, Y_t . Let $F(x_t|x_{t-1}^{(k)}, y_{t-1}^{(l)})$ denote the distribution function of the target variable X , conditional on the joint (k, l) history $X_{t-1}^{(k)}, Y_{t-1}^{(l)}$. Then, variable Y is said to be Granger-cause variable X (with lags k, l) if and only if [42, 44]

$$F(x_t|x_{t-1}^{(k)}, y_{t-1}^{(l)}) \neq F(x_t|x_{t-1}^{(k)}). \tag{3}$$

Thereby, it is said that Y Granger-causes X if and only if X is not independent of the history of Y .

There exists a series of results [31, 34, 45] that state an exact equivalence between the Granger causality and TE statistics for various approaches and assumptions on the data generating processes, which make it possible to construct TE as a nonparametric test of pure Granger causality. This connection can be regarded as a bridge between causal inference of data under autoregressive models and the information-theoretic approach. Before proceeding, we want to emphasize that for highly nonlinear and non-Gaussian data, as is the case for many financial instruments, it is better to approach causality by using the TE information method instead of the traditional Granger causality test [44].

There are several techniques for estimating TE from observed data in order to apply it to real-world data problems. However, most of them require a large amount of data, and consequently, their results are commonly biased due to small-sample effects, which limits the use of TE in practical data applications. To avoid this problem, we use the robust and computationally fast technique of symbolization [38] to estimate TE. Symbolic transfer entropy (STE) has been introduced within the concept of permutation entropy [46]. Following [38, 46], symbols

are defined by reordering the amplitude values of time series x_i and y_i . Thus, for a given i, m arbitrary amplitude values, the elements of the delay vector

$$\{x(i), x(i + l), \dots, x(i + (m - 1)l)\}, \tag{4}$$

are arranged in ascending order

$$\{x(i + (k_{i1} - 1)l) \leq x(i + (k_{i2} - 1)l) \leq \dots \leq x(i + (k_{im} - 1)l)\}, \tag{5}$$

where l denotes the time delay, i.e., the time between successive observations of $x(i)$; and m represents the embedding dimension, which is interpreted in the time series context as the total number of past data points in the delay vector. A symbol is thus defined as $\hat{x}_i = (k_{i1}, k_{i2}, \dots, k_{im})$, and the relative frequency of symbols is used to estimate the joint and conditional probabilities of the sequence of permutation indices. One of the advantages of STE is the delay vector formed by time series segments can be compared in terms of rank. Thus, similar data patterns can be found regardless of their magnitude levels, extending its validity to non-stationary systems [39].

To provide an example of this procedure, let us consider the time series {1, 2, 3, 6, 5, 4} to estimate the related Shannon entropy [28] measure for $l = 1, m = 2$. First, we need to organize the five pairs of adjacent figures according to their relative values. As a result, three pairs are observed to satisfy the relation $x_t < x_{t+1}$ characterized by permutation {01}, and two pairs for which $x_t > x_{t+1}$ represent permutation {10}. Then, the Shannon entropy for $l = 1, m = 2$ is given by

$$H(2) = -(3/5) \log(3/5) - (2/5) \log(2/5) \approx 0.971. \tag{6}$$

Let us now return to the original problem of TE estimation. Given symbol sequences $\{\hat{x}_i\}$ and $\{\hat{y}_i\}$, STE is mathematically defined as [38]

$$T_{Y \rightarrow X}^S = \sum_{ij} p(\hat{x}_{i+\delta}, \hat{x}_i, \hat{y}_i) \log \frac{p(\hat{x}_{i+\delta} | \hat{x}_i, \hat{y}_i)}{p(\hat{x}_{i+\delta} | \hat{x}_i)}, \tag{7}$$

where the sum is over all symbols, and δ denotes a time step. The logarithm is with base 2; thus, $T_{Y \rightarrow X}^S$ is in bits.

The question at this point is whether a given empirical measurement of STE is statistically different from 0 and represents sufficient evidence of a direct relationship between the variables. It is possible to construct a null hypothesis H_0 that there is no such relationship, but it is necessary to know what the distribution of an empirical measurement would look like if H_0 were true and then evaluate a *p-value* for sampling the actual measurement from the distribution. If the test fails, we will accept the alternate hypothesis that there exists a direct relationship.

For discrete X and Y , it is known that if H_0 is true, then $T_{Y_s \rightarrow X}^S \xrightarrow{d} \chi^2(d)/(2N \log 2)$, where the number of degrees of freedom d is the difference between the number of parameters in the full and null models [31]. Y_s represents surrogate variables for Y generated under H_0 , which have the same statistical properties as does Y , but any potential correlation with X is destroyed. As a consequence, surrogates of the distribution $T_{Y_s \rightarrow X}^S$ must preserve $p(\hat{x}_{i+\delta} | \hat{x}_i)$ but not $p(\hat{x}_{i+\delta} | \hat{x}_i, \hat{y}_i)$ [47].

In order to evaluate the advantage of STE over the Granger causality approach let us consider the following simulated systems:

1. A three-variable system, with nonlinear drivers

$$\begin{aligned}
 x_1(t) &= 0.7x_1(t-1) + \varepsilon_1(t) \\
 x_2(t) &= 0.3x_2(t-1) + 0.5x_2(t-2)x_1(t-1) + \varepsilon_2(t) \\
 x_3(t) &= 0.3x_3(t-1) + 0.5x_3(t-2)x_1(t-1) + \varepsilon_3(t),
 \end{aligned}
 \tag{8}$$

where $\varepsilon_i \sim N(0, 1)$.

The term product of the variables in the second and third equation causes the variables x_2 and x_3 to have marginal distributions with long tails [39]. Therefore this model characterizes the heavy-tail behavior of financial time series.

2. A ten-variable system with overlapping linear drivers separated in two blocks

$$x_i(t) = \sum_j^{10} A_{ij}x_j(t-1) + \varepsilon_i(t), \quad i = 1, \dots, 10;
 \tag{9}$$

where $\varepsilon_i \sim N(0, 1)$, and the interaction matrix A has the form

$$A = \begin{pmatrix}
 a & b & c & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & a & b & c & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & a & b & c & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & a & b & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & a & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & a & b & c & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & a & b & c & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & b & c \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & b \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a
 \end{pmatrix}
 \tag{10}$$

where we set $a = 0.9$, $b = 0.7$, and $c = 0.3$.

In the simulations of each system, a time series length of dimension $T = 2048$ is considered after removed the transitory effect of 256 steps. For STE the time delay is set to $l = 1$ as usual practice, and the embedding dimension is set to $m = 2$ for both systems. The embedding specification is because the maximum lag is two for *system 1*, as well as m must be bigger than one by definition. Likewise, the order P of the VAR model associated to the Granger causality test is set to $P = 2$ for *system 1*, and $P = 1$ for *system 2* in accordance with the lag order of each model. As a criterion of a causal relation, the null hypothesis is rejected at the level of *p-value* = 0.01. In the Granger approach has been considered the F-distribution and the chi-square distribution to test the coefficients of the causal time series. Thus, the null of zero value coefficients must be rejected in both tests in order to consider a true causal relation, and the F-value is used as the quantity to compare the magnitude of the causal relationships. Further, the visual representation is done via a directed graph $G = (V, E)$, where the nodes V represent the involved variables, and the direction of the edges V represent the causal relations given by the

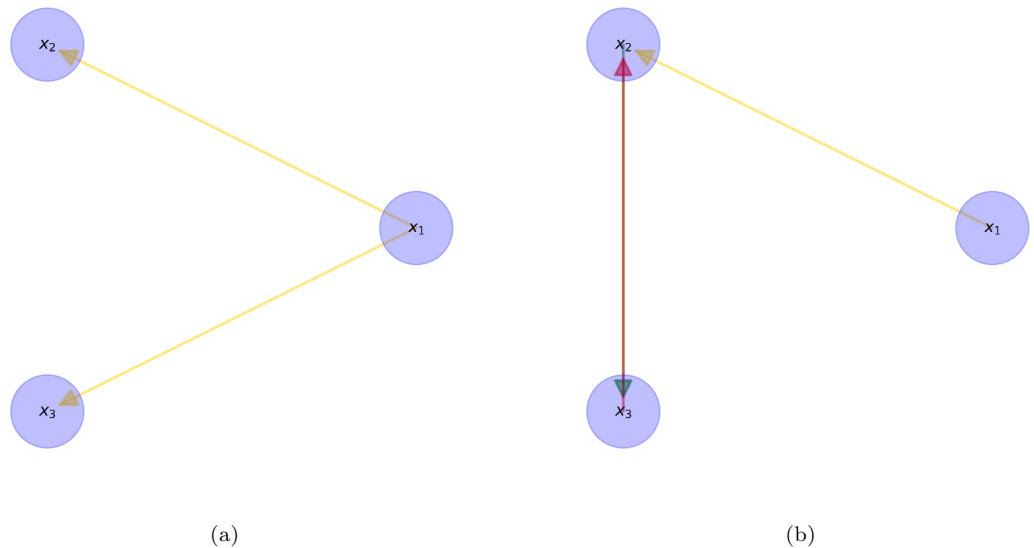


Fig 1. System 1. (a) Represents the STE detected interactions (b) Represents the Granger causality detected interactions.

<https://doi.org/10.1371/journal.pone.0227269.g001>

significant STE or F values, for the transfer entropy and Granger causality approach, respectively.

In *system 1* it is modeling a system with nonlinear drivers $x_1 \rightarrow x_2$ and $x_1 \rightarrow x_3$, where the variables x_2 and x_3 comes from a distribution with long-tails. In Fig 1 it is shown the associated directed graph for (a) STE and (b) Granger causality measures, respectively. We can notice STE correctly detects the nonlinear direct causality, whereas Granger causality is not able to describe the two nonlinear interactions, but on the contrary indicates two spurious causal effects $x_2 \rightarrow x_3$ and $x_3 \rightarrow x_2$.

In *system 2*, it is modeling a system where each variable is linearly driver by at most two variables, and the interactions as a whole are separated into two groups. The linear interaction matrix A is shown in Fig 2a as a heatmap, and in Fig 2b as a directed graph. In the same way, Fig 2(c), 2(d), 2(e) and 2(f) show the detected causal relations by STE and Granger causality, respectively. We can notice STE only detects causal effects within the cluster, while Granger causality on the contrary also detects spurious interactions between clusters. Nevertheless both approaches detect spurious relations within clusters, the number of false-positive interactions is less for STE than for Granger causality, with 10 and 21, respectively. On the other hand, the number of false-negatives is two for STE, whereas Granger causality is zero. In Fig 2c and 2e is shown the scaled magnitude of STE and Granger Causality in values between zero and one, in this scaled representation, the magnitude of the false-positives interactions in the Granger approach is 33% more than in the STE approach.

In the last models, we have tested linear and nonlinear causal interactions. In the nonlinear model, the Granger causality approach does not was able to detect the interactions and even indicates two more spurious drivers, while STE correctly describes the system. In the linear model, both approaches detects spurious drivers, but the number and magnitude were bigger under the Granger measure. Also, it does not discriminate against the cluster structure of the system. Despite the fact we have tested causalities in bivariate form, STE can capture the main characteristics of the system, which allows us to rely on its use for the applications as the study we are interested in below.

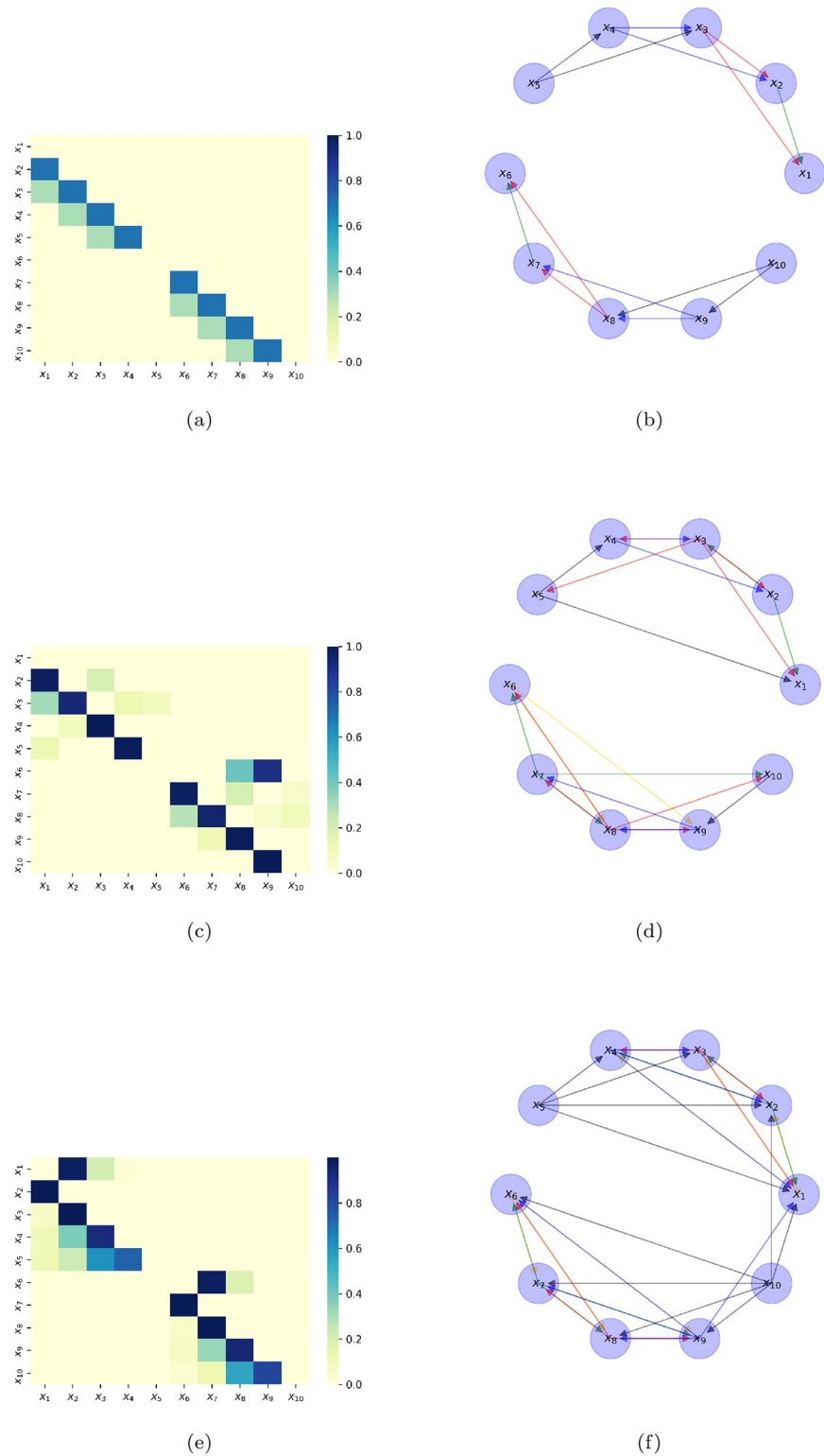


Fig 2. System 2. (a) Heatmap of the interaction matrix A. (b) Directed graph of the interaction matrix A. (c) Heatmap of the STE detected interactions. (d) Directed graph of the STE detected interactions. (e) Heatmap of the Granger causality detected interactions. (f) Directed graph of the Granger causality detected interactions.

<https://doi.org/10.1371/journal.pone.0227269.g002>

Table 1. STE results.

Δt	m	$\sum_{ab} T_{ab}^s$	$\#\{T_{ab}^s > 0\}$
0	2	7.7484	4221
0	3	97.7024	6345
0	4	241.6957	736
1	2	19.677	7756
1	3	118.1083	7067
1	4	351.52	1069
2	2	1.3937	1289
2	3	68.196	4701
2	4	442.0707	1342
3	2	1.3346	1240
3	3	13.8508	1070
3	4	333.1614	1013

STE of cryptocurrency return time series for each of lag times $\Delta t = 0, 1, 2, 3$ for the pair of predictor-response variables X, Y and embedding dimension $m = 2, 3, 4$. The third column shows the total amount of direct information at the p -value of 0.10, while the fourth column shows the corresponding number of preserved relations at the same level of statistical significance.

<https://doi.org/10.1371/journal.pone.0227269.t001>

Application to cryptocurrency data

To present our results in the context of a predictive model, let us refer to variable $x(t)$ as the predictor and variable $y(t)$ as the response. Thus, we estimate STE for the combination of pairs $\{X_a(t), Y_b(t + \Delta t)\}$, where $a, b = 1, \dots, p$ ($= 100$); $t = 0, \dots, n - \Delta t$, and Δt is a time lag added to consider predictive scenarios. The STE estimation needs the specification of the time delay l and the embedding dimension m as has been explained above. In this study l is fixed and m is varied as usual practice to simplify the state-space reconstruction [43]. The results for time delay $l = 1$ and p -value = 0.10 are given in Table 1 for various values of lag time Δt and embedding dimension m . The third column shows the total sum of $T_{X_a \rightarrow Y_b}^s$ for all possible combinations of indices a, b , as long as a direct relationship exists under H_0 . The fourth column shows the number of relations that are preserved. We observe a peak in the number of preserved relations at $\Delta t = 1$ and $m = 2, 3$, where the number of relations exceeds 7000 out of 10000 possible relations. Even though the maximum is reached at $m = 2$, we chose the case $m = 3$ following the criterion of obtaining at the same time the maximum of the total sum of information flow (118.1084 bits).

Moreover, we show in Fig 3a and 3b the heat map of STE results for $m = 2$ and $m = 3$, respectively. It is observed that STE has higher values in Fig 3b than in Fig 3a in general. Further, some structure can be noticed in the upper left section of Fig 3a, which is sharper in Fig 3b. This upper left section corresponds to cryptocurrencies with the highest capitalization due to the way we order them. Therefore, it is natural to have the highest values of information flow in that sector.

A convenient procedure for measuring the net flow of information between processes X and Y is by using the normalized directionality index (NDI), given by [44]

$$d(X, Y) = \frac{STE_{X \rightarrow Y} - STE_{Y \rightarrow X}}{STE_{X \rightarrow Y} + STE_{Y \rightarrow X}} \in [-1, 1]. \tag{11}$$

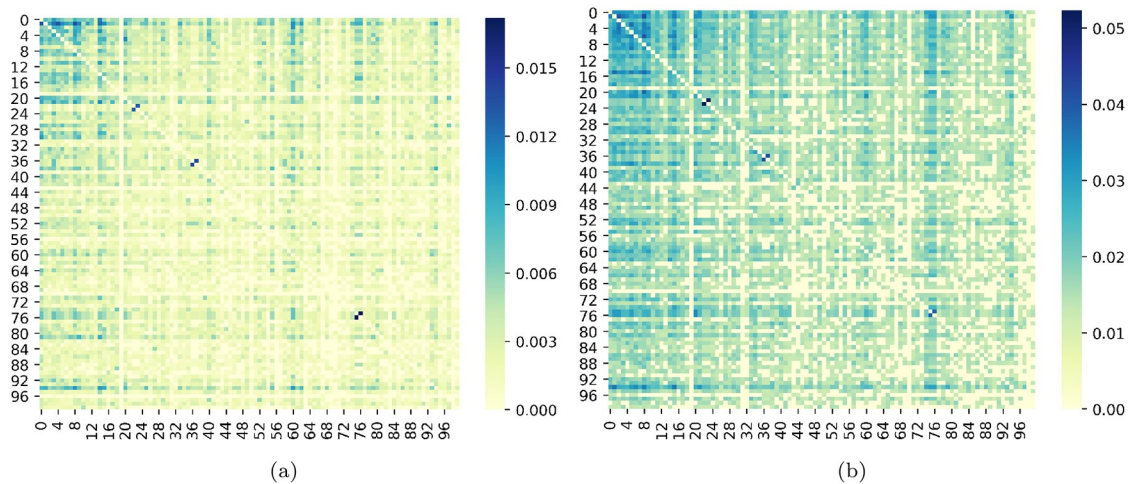


Fig 3. Heat map of STE. (a) $m = 2$. (b) $m = 3$. The color intensity represents the magnitude of STE.

<https://doi.org/10.1371/journal.pone.0227269.g003>

This quantity standardizes *STE* values to be between -1 and 1 , so $d(X, Y)$ is maximized if one of the *STE* values is zero and minimized if they are equal. This index is not normalized in the statistical sense but resembles a measure of divergence or market leverage and, beyond that, is very useful in comparing measures across different systems or financial sectors. The form of Eq (11) gives importance to the relative information flows between X and Y enabling compare different relations in the same scale. This feature help us to consider as a drivers of Y , the X 's variables for which $d(X, Y) > 0$. In this way we set the preferred direction of information flow and avoid the ambiguity of having both directions with different weights, which help to compute the variable selection in straightforward manner by simple concepts of directed graph.

We apply NDI to our previous results for $\Delta t = 1$ and $m = 3$. Again, for a better visualization, the obtained values are first converted to a directed graph $G = (V, E)$, where the nodes V represent cryptocurrencies, and the edges E now represent the positive values of $d(X, Y)$ resulting from applying NDI. Fig 4 shows as an example a directed subgraph for the first 10 cryptocurrencies in the order of capitalization with its corresponding edges given by the NDI measure. In this figure, the arrow direction indicates the information flows from one variable to another. We observe, e.g., that the *eos* coin only receives information from the other cryptocurrencies under the NDI measure, whereas the *ripple* sends and receives information from the members of the subnet.

To discriminate the predictor variables from the response variables, several basic concepts of graph theory are used. The node out-degree k^{out} is the number of edges pointing out from the node, while the node in-degree k^{in} is the number of edges pointing towards the node. We use these concepts to select the sets of predictor-response variables by the proposed heuristic selection rule:

$$\begin{aligned}
 V_i \in \{\text{response variables}\} & \quad \text{if } k_i^{in} \geq k_i^{out}, \\
 V_i \in \{\text{predictor variables}\} & \quad \text{if } k_i^{in} < k_i^{out},
 \end{aligned}
 \tag{12}$$

for $i = 1, \dots, p$. The results of applying this procedure are shown in Table 2 for the first 10 response and predictor variables (see S2 Table for the entire list). In general, we found 49

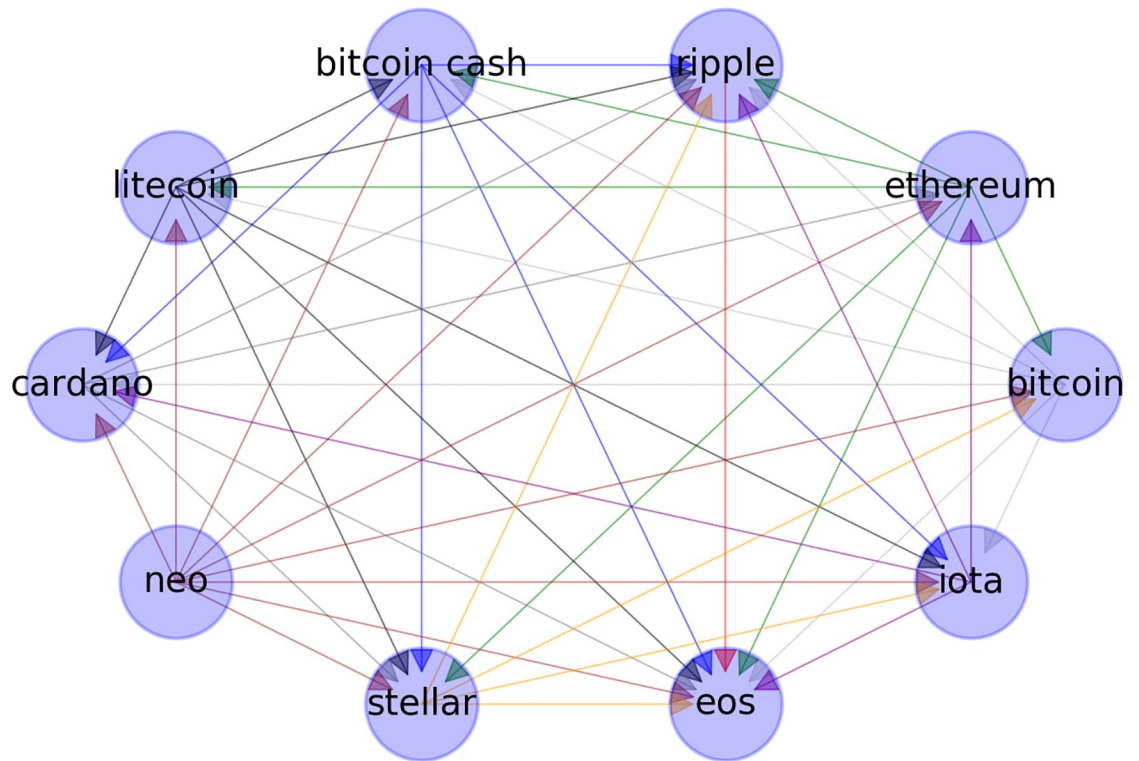


Fig 4. NDI subgraph. The arrow direction represents the direction of the information flow.

<https://doi.org/10.1371/journal.pone.0227269.g004>

predictor variables and 51 response variables in our set of $p = 100$ times series of cryptocurrency returns.

Modularity and centrality

To confirm that this variable selection makes sense, we have found the structure of the associated clusters through the modularity approach. This technique is based on the principle that

Table 2. Predictor and response variables.

i	Predictor (49)	Response (51)
1	ethereum	bitcoin
2	neo	ripple
3	dash	bitcoin cash
4	monero	litecoin
5	lisk	cardano
6	bitcoin gold	stellar
7	tether	eos
8	steem	iota
9	populous	nem
10	siacoin	ethereum classic
⋮	⋮	⋮

The first 10 predictor and response variables in order of capitalization, selected using the heuristic criterion given above. The total number of selected variables is shown in parentheses.

<https://doi.org/10.1371/journal.pone.0227269.t002>

networks that tend to form clusters differ from random networks [48]. It measures the deviation from a random network for a specific partition of the nodes into clusters. Then, we find the optimal partition maximizing the modularity version for directed networks [49] defined by

$$Q = \frac{1}{N_E} \sum_{i,j} \left[A_{ij} - \frac{k_i^{out} k_j^{in}}{N_E} \delta(c_i, c_j) \right], \tag{13}$$

where A_{ij} is the binary entry of the adjacency matrix, which represents the existence or not of an edge between nodes i and j by one or zero, respectively; c_i is the label of the cluster to which node i is assigned to, and N_E is the total number of edges in the network. Then, the implicit idea behind is to maximize Q over possible divisions of the network into clusters or communities.

In the top graphs of Fig 5 it is shown the generated clusters via our heuristic approach (Fig 5a) and the optimal generated clusters under the modularity approach (Fig 5b), both results for the case $\Delta t = 1, m = 3$. In spite of the fact that under the modularity approach the NDI network form three clusters, they resemble our predictor and response sets. It is worth to mention that the visualization is made using the Force-directed algorithm [50]. Here, the position of each node is fixed by considering the weights of the edges as an attractive spring force and is seek equilibrium of forces in each node by adding an electrostatic repulsive force as a free parameter. In addition, we have separated the clusters from its equilibrium position for visualization purposes by a radial distance of $\sqrt{1/2}$ to the top-right direction if the node belongs to the response set, and to the same distance but to the bottom left direction if the node belongs to the predictor set under. In this manner, the node position does not change in both results since the algorithm only considers the associated weights to fix the position. On the contrary, the color changes according to the cluster at which each node corresponds.

A surprising result is that *ripple* is consistently assigned by modularity to a cluster with a similar pattern as the response set. This behavior is interesting because ripple looks quite different from all other cryptocurrencies in the force-directed visualization. A possible explanation is due to the particular characteristics of *ripple* as a digital payment protocol and its increasing adoption by banks for its faster international transactions over the traditional swift payment method.

In the middle graphs of Fig 5 it is shown the equivalent analysis of above for the case $\Delta t = 1, m = 2$, where we can see that the variable selection nodes via our heuristic rule are not completely separated in the sense of forces (Fig 5c), but in this case the modularity approach found two clusters (Fig 5d). To contrast this behavior, in the bottom graphs of Fig 5 we have shown one of the worst cases under our heuristic rule, corresponding to $\Delta t = 0, m = 4$, where $\#\{T_{ab}^s > 0\}$ has the lower value in Table 1. Here we can see that the external force of $\sqrt{2}$ applied to separate the sets or response and predictor nodes is stronger than the internal force between the nodes, as a consequence, a pattern is more difficult to find. This can be confirmed by the modularity approach, which has found five clusters in this case.

In Table 3 we show the percentage of identification of each cluster found by modularity in relation to the predictor and response variable sets categorized under our selection rule. It can be seen case $\Delta t = 1, m = 2$ has the best identification, where the elements of cluster one given by modularity coincide with the elements of the predictor set in an 87.5%, while cluster two coincide with the response set in a 90.38%. Similarly, our preferred case of $\Delta t = 1, m = 3$ have a coincide of 83.67% and 82.35%, respectively. Nevertheless the case $\Delta t = 1, m = 2$ have higher coincide percentage, we still prefer the case $\Delta t = 1, m = 3$ over it as the best partition because the larger separation of the sets under the force-directed algorithm (Fig 5a and 5b).

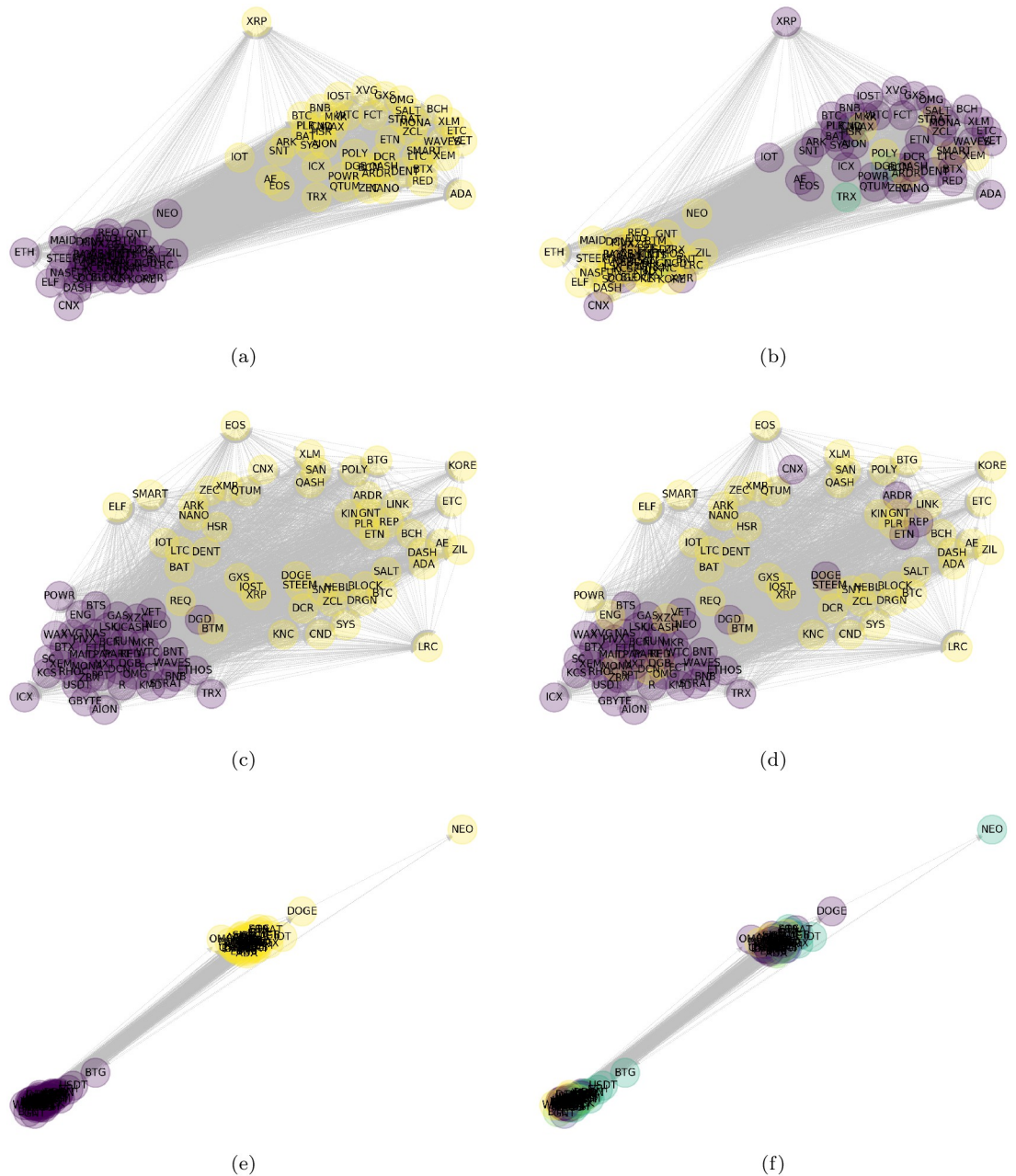


Fig 5. Predictor-reponse clusters vs. modularity clusters. Comparison of the cluster formed by the heuristic variable selection rule of Eq 12 (left figures) and the cluster formed by the modularity approach of Eq 13 (right figures). From top to bottom the cases $\Delta t = 1, m = 3$ ((a) and (b)), $\Delta t = 1, m = 2$ ((c) and (d)), $\Delta t = 0, m = 4$ ((e) and (f)). The node names represent the symbol of the associated cryptocurrency as listed on the annexed table in the appendix (see S1 Table).

<https://doi.org/10.1371/journal.pone.0227269.g005>

To explore further the structure of the directed network associated with NDI of the transfer entropy, we have computed the in-degree and out-degree centralities of the whole network. The in-degree centrality for a node V is defined as the fraction of nodes in which incoming edges are connected to. The out-degree centrality for a node V is the fraction of nodes whose outgoing edges are connected to. Before showing the results, it is important to mention that cryptocurrencies are divided into two main categories: coins and tokens. Basically, coins have

Table 3. Cluster identification.

$\Delta t = 1, m = 3$	Predictor (49)	Response (51)
Cluster 1	83.67%	11.76%
Cluster 2	12.25%	82.35%
Cluster 3	4.08%	5.88%
$\Delta t = 1, m = 2$	Predictor (48)	Response (52)
Cluster 1	87.5%	9.61%
Cluster 2	12.5%	90.38%
$\Delta t = 0, m = 4$	Predictor (56)	Response (44)
Cluster 1	14.29%	22.73%
Cluster 2	28.57%	31.82%
Cluster 3	25%	15.91%
Cluster 4	21.43%	20.46%
Cluster 5	10.71%	9.09%

Percentage of coincidence between the predictor-response sets and the clusters found by the modularity approach for different values of Δt and m .

<https://doi.org/10.1371/journal.pone.0227269.t003>

the same characteristics as money. i.e., they are fungible, divisible, acceptable, portable, durable and have limited supply, while tokens are issued by the project, which can be used as a method of payment inside the project’s ecosystem, in this way they are more like digital assets. Tokens, unlike coins, are created on top of existing blockchains. By far the most common platform to create tokens is the Ethereum platform. Other token platforms include Stellar, NEO, Omni, and EOS.

In [Table 4](#) it is shown the top ten in-degree and out-degree centrality ranking of cryptocurrencies, from the highest to the lowest. The first and fourth column represents the order number in terms of capitalization, the second and fifth the name of the cryptocurrency, the third and sixth the type, i.e., coin or token, and finally, the seventh column specify from which platform was created the token. The results show that the highest in-degree centrality values are dominated by coins, while mainly tokens have the highest out-degree values. This result tells us that coins are central in the sense of receiving information, while tokens are central in the sense of sending information. Hereof, the direction of information transfer helps to discriminate between tokens and coins.

Table 4. In-degree and out-degree centrality.

in-degree			out-degree			
number	name	type	number	name	type	platform
2	ripple	coin	42	komodo	coin	
8	eos	coin	62	kyber network	token	ethereum
14	ethereum classic	coin	81	funfair	token	ethereum
5	cardano	coin	52	digixdao	token	ethereum
3	bitcoin cash	coin	72	ethos	token	ethereum
20	icon	coin	94	bancor	token	ethereum
9	iota	coin	13	lisk	token	ethereum
23	raiblocks	coin	53	gas	token	NEO
91	reddcoin	coin	38	0x	token	ethereum
22	zcash	coin	18	bitcoin gold	coin	

Top ten in-degree and out-degree centrality ranking of cryptocurrencies.

<https://doi.org/10.1371/journal.pone.0227269.t004>

Table 5. Hubs and authorities.

Hubs				Authorities			
number	name	type	platform	number	name	type	platform
98	particl	coin		91	reddcoin	coin	
82	kin	coin		63	monacooin	coin	
95	tenx	token	ethereum	80	dent	token	ethereum
90	chainlink	token	ethereum	2	ripple	coin	
87	neblio	coin		73	pillar	token	ethereum
67	dentacoin	token	ethereum	23	raiblocks	coin	
50	revain	token	ethereum	77	factom	coin	
42	komodo	coin		48	zclassic	coin	
97	santiment	token	ethereum	22	zcash	coin	
54	byteball	coin		14	ethereum_classic	coin	

Top ten Hub and Authority centrality ranking of cryptocurrencies.

<https://doi.org/10.1371/journal.pone.0227269.t005>

In some networks, it is appropriate also to accord a vertex high centrality if it points to others with high centrality. There are two types of important nodes in this context: authorities and hubs. Authorities are nodes that contain useful information on a topic of interest; hubs are nodes that tell us where the best authorities are to be found [51]. Specifically, authority centrality is defined as the sum of the hub centralities y_j which point to the node x_i

$$x_i = \alpha \sum_j^p A_{ij} y_j, \tag{14}$$

for $i = 1, \dots, p$, where α is constant. Likewise, Hub Centrality is the sum of the authorities x_j which point to the node y_i

$$y_i = \beta \sum_j^p A_{ij} x_j, \tag{15}$$

for $i = 1, \dots, p$, with constant β . In matrix notation

$$\mathbf{x} = \alpha \mathbf{A} \mathbf{y}, \quad \mathbf{y} = \beta \mathbf{A}' \mathbf{x} \tag{16}$$

These expressions were proposed and developed by Kleinberg [52] into a centrality algorithm called hyperlink-induced topic search or HITS.

In Table 5 it is shown the top ten hub and authority centrality ranking of cryptocurrencies, from the highest to the lowest. The first and fifth column represents the order number in terms of capitalization, the second and sixth the name of the cryptocurrency, the third and seventh the type, i.e. coin or token, and finally, the fourth and eighth column specify from which platform was created the token. The results show that there is not a clear pattern in the hub centrality values, but authorities are dominated by coins. This can be interpreted as coins are central authorities of the transfer entropy network in the sense that they are pointed in by a lot of good hub cryptocurrencies. This means that coins are authorities in the ecosystem of cryptocurrencies.

Regression framework

Now that the set of predictor-response variables has been determined and analyzed, we would like to present a general regression model as the framework to forecast the response variables.

Thus, this section describes the model and the related problem of rank determination that highlights the need to study some results from random matrices.

Consider the reduced-rank regression (RRR) model given by [53]

$$Y = \mu + CX + \epsilon \tag{17}$$

where μ and C are unknown regression parameters, and the unobservable error variate ϵ of the model has mean $E(\epsilon) = 0$ and covariance matrix $cov(\epsilon) = E\{\epsilon\epsilon'\} = \Sigma_{\epsilon\epsilon}$, and is distributed independently of X . The difference from the classical multivariate regression model is that the rank of the regression coefficient matrix C is deficient

$$rank(C) = \eta \leq \min(r, s). \tag{18}$$

The rank condition implies that there may be a number of linear constraints on the set of regression coefficients in the model. Given a sample X, Y of observations, the goal is to estimate optimally the parameters μ and C . Hence, the idea is to minimize the objective function

$$W(t) = E\{(Y - \mu - CX)' \Gamma (Y - \mu - CX)\}, \tag{19}$$

where Γ is a positive definite symmetric matrix of weights, and the expectation is taken over the joint distribution of X, Y .

RRR can be regarded as a unifying treatment of several classical multivariate procedures that were developed separately from each other. If we set X (and $r = s$) by making the output variables identical to the input variables and, in addition, set $\Gamma = I$, then we obtain Harold Hotelling’s PCA and exploratory factor analysis. If we set $\Gamma = \Sigma_{YY}^{-1}$, then we have Hotelling’s canonical variance and correlation analysis. A nonlinear generalization of RRR provides a flexible model for artificial neural networks [54]. Nevertheless, one of the primary and most difficult parts of model determination is to assess the unknown value of parameter η , called the effective dimensionality of the multivariate regression. The reduction in $W_{min}(\eta)$ obtained by increasing the rank from $\eta = \eta_0$ to $\eta = \eta_1$, where $\eta_0 < \eta_1$, is given by

$$W_{min}(\eta_0) - W_{min}(\eta_1) = \sum_{j=\eta_0+1}^{\eta_1} \lambda_j. \tag{20}$$

This relation depends upon Γ only through the eigenvalues $\{\lambda_j\}$ of

$$N = \Gamma \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \Gamma \tag{21}$$

However, the value of η and, hence, the number and nature of those constraints may not be known prior to statistical analysis.

We are interested in estimate the parameter η for the particular case of $\Gamma = \Sigma_{YY}^{-1}$, i.e., the canonical correlation analysis, and explore the properties and behavior of the number of factors through time. Our focus is on the high-dimensional settings under the random matrices approach. Hence, in the next section, we introduce the theory step-by-step and describe the results for the predictor-response sets of cryptocurrencies. The core of the technique is the deep connection of the estimation of η via the Tracy-Widom distribution as we will see in what follows.

Number of factors

The random matrix theory (RMT) is an important framework for analyzing limit distributions of eigenvalues. Historically, RMT was developed to solve complex problems in nuclear physics

and more recently in quantum chaos [55]. During the preceding decades, seminal applications of RMT arose in the context of mesoscopic physics, biological microarrays, wireless communication and econophysics [56–60]. A common ingredient of the cited studies is the following result, restated here using the terminology of high-dimensional statistics.

Let X be a $p \times n$ matrix, where the elements $X_{i,j}$ are i.i.d. random variables with distribution $N(0, 1)$. Then, as $p, n \rightarrow \infty$, such that $\frac{n}{p} \rightarrow c \in (0, \infty)$, the spectral density of the Wishart matrix $W = n^{-1}XX'$ converges (a.s.) to the Marcenko-Pastur law [25]

$$\rho(\lambda) = \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{2\pi c\lambda}, \tag{22}$$

where

$$\lambda_{min}^{max} = (1 \pm \sqrt{c})^2. \tag{23}$$

In the econophysics community, the Marchenko-Pastur distribution is known as a universal result of Wishart matrices. If there is no correlation between financial variables, then the eigenvalues of their correlation matrix should be bounded by this RMT prediction [59, 60].

In statistics, it is of primary importance to consider null hypothesis tests. The Wishart matrices that appear in the preceding result can be denoted by $W_p(n, I)$, where I is the covariance matrix of the population distribution of $n^{-1}XX'$. In our case, it is of interest to test the hypothesis of identity covariance matrix, $H_0: \Sigma = I$, against an alternative case $H_A: \Sigma \neq I$, where Σ has some more general structure. Using this approach, it is possible to compute a confidence interval for accepting or rejecting the universal result of Wishart matrices of empirical datasets for the general range of dimensions p and n . The approach to quantifying a confidence level is based on the approximation to the null hypothesis' distribution of the largest sample eigenvalue $\hat{\lambda}_1$

$$P\{\hat{\lambda}_1 > M : W \sim W_p(n, I)\}, \tag{24}$$

which is the probability of finding $\hat{\lambda}$ larger than an upper bound M , given that $W = n^{-1}XX'$ follows the Wishart distribution $W_p(n, I)$. The following result of the random matrix theory leads to the needed approximate distribution [10].

Assume $A \sim W_p(n, I)$, $p/n \rightarrow \gamma \in (0, \infty)$, and denote by $\hat{\lambda}_1$ the largest eigenvalue in the eigenvalue equation $Au = \hat{\lambda}u$. Then, the distribution of the largest eigenvalue approaches that of the Tracy–Widom F_β law

$$P\{n\hat{\lambda}_1 \leq \mu_{np} + \sigma_{np}s | H_0\} \rightarrow F_\beta(s) \tag{25}$$

where $\mu_{np} = (\sqrt{n} + \sqrt{p})^2$, $\sigma_{np} = \mu_{np} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}}\right)^{1/3}$. There are elegant formulas for computing the Tracy-Widom distribution functions:

$$\begin{aligned} F_1(s) &= \sqrt{F_2(s) \exp\left(-\int_s^\infty q(x)dx\right)} \\ F_2(s) &= \exp\left(-\int_s^\infty (x-s)^2 q(x)dx\right), \end{aligned} \tag{26}$$

in terms of the solution q of a nonlinear second-order differential equation $q'' = sq + 2q^3$, $q(s) \sim A_i(s)$ as $s \rightarrow \infty$, also known as the classical Painlevé type II equation. Functions in the family F_β are obtained numerically as a function of q . Even though some effort is required to

solve for F_β , from the point of view of applied data analysis, they are special functions, as is the normal curve [61].

Let us provide an example of the relevance of the Tracy-Widom test. Suppose that in a sample of $n = 10$ observations from a $p = 10$ -dimensional Gaussian distribution $N_{10}(0, \Sigma)$, the largest sample eigenvalue $\lambda_1 = 4.25$ has been obtained. Given these dimensions, the support of the Marchenko-Pastur distribution is bounded to the interval $[0, 4]$ (see Eq (23)) Then, the question in statistical terms is whether an observed largest eigenvalue of 4.25 is consistent with $H_0: \Sigma = I$ if $n = p = 10$. The second-order Tracy-Widom approximation [62] yields 6% chance of observing a value more extreme than 4.25 even if no structure is present, i.e., $\Sigma = I$. Considering the traditional 5% benchmark, this is not strong enough evidence to reject the null hypothesis H_0 [63].

The Tracy-Widom test becomes relevant to the determination of the number of components that must be retained in PCA, especially in the context of high-dimensional data, i.e., if $\mathcal{O}(n/p) = 1$. Beyond PCA, there are several classical problems in multivariate statistics that can take advantage of this type of test. These problems can be generalized under the greatest root distribution. It describes the null hypothesis of apparently different problems, including multiple-response linear regression, multivariate analysis of variance, canonical correlations, equality of covariance matrices, etc. [64]. The next definition from [65] states the greatest root distribution formally.

Let $\mathbf{A} \sim W_p(l, \mathbf{I})$ be independent of $\mathbf{B} \sim W_p(n, \mathbf{I})$, where $l \geq p$. Then, the largest eigenvalue θ of $(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$ is called the greatest root statistic, and its distribution is denoted by $\theta(p, l, n)$. It has the property

$$\theta(p, l, n) \stackrel{d}{=} \theta(n, l + n - p, p), \tag{27}$$

that is useful, in particular, if $n < p$.

There is an interesting connection between the greatest root statistic and Tracy-Widom distributions. In the study of Johnstone [13], it is shown that with appropriate centering and scaling, the distribution of the logit transform W of θ approximates a Tracy-Widom distribution:

$$\frac{W(p, l, n) - \mu(p, l, n)}{\sigma(p, l, n)} \xrightarrow{d} F_1, \tag{28}$$

where

$$W(p, l, n) = \text{logit}\theta(p, l, n) = \log \left(\frac{\theta(p, l, n)}{1 - \theta(p, l, n)} \right) \tag{29}$$

is the logit transform of θ , and the centering and scaling parameters are defined by

$$\mu(p, l, n) = 2 \log \tan \left(\frac{\phi + \gamma}{2} \right), \sigma^3(p, l, n) = \frac{16}{(l + n - 1)^2} \frac{1}{\sin^2(\phi + \gamma) \sin \phi \sin \gamma}, \tag{30}$$

where the angle parameters γ, ϕ are defined as

$$\sin^2 \left(\frac{\gamma}{2} \right) = \frac{\min(p, n) - 1/2}{l + n - 1}, \sin^2 \left(\frac{\phi}{2} \right) = \frac{\max(p, n) - 1/2}{l + n - 1}. \tag{31}$$

At this point, we are interested in describing a procedure for determining parameter η in the RRR model by using the greatest root statistic, which will reconcile both frameworks. This is traditionally accomplished by the canonical correlation analysis (CCA). The respective approach involves partitioning a collection of variables into two sets. Consider a set \mathbf{X} with q variables and a set \mathbf{Y} with p variables. The objective is to find maximally correlated

combinations $\eta = \mathbf{a}'\mathbf{x}$ and $\phi = \mathbf{b}'\mathbf{y}$. Even though CCA has maximal properties similar to those of PCA, the objective of canonical correlation concerns the relationship between two groups of variables instead of interrelationships within a set of variables.

Suppose that (\mathbf{X}, \mathbf{Y}) is a data matrix of n observations on $q + p$ variables such that each sample is independent of the others and has the populations distribution $N_{p+q}(\mu, \Sigma)$. Assume the sample covariance matrix \mathbf{S} partitioning

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \mathbf{S}_{YX} & \mathbf{S}_{YY} \end{pmatrix}. \tag{32}$$

The sample squared canonical correlations (r_i^2) for $i = 1, \dots, k = \min(p, q)$ are obtained as the eigenvalues of $\mathbf{M}_S = \mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}$, whereas the population counterparts are given by the eigenvalues of $\mathbf{M}_\Sigma = \Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$ [65]. Note that the nonzero eigenvalues of \mathbf{M}_Σ are the same as the nonzero eigenvalues of \mathbf{N} in Eq (21) for $\Gamma = \Sigma_{YY}^{-1}$, which is precisely the CCA case in the general RRR model.

We are now interested in describing the procedure of testing the null hypothesis of independence of the two sets of variables $H_0: \Sigma_{12} = 0$ by using the Tracy-Widom test. First, let us point out the next result concerning joint independence of partitioned Wishart matrices.

Let $\mathbf{M} \sim W_p(n, \Sigma)$, and partition matrix \mathbf{M} into submatrices \mathbf{M}_{11} of dimensions $a \times a$ and \mathbf{M}_{22} of dimensions $b \times b$, where $a + b = p$ and $n > a$. Define the product of matrices $\mathbf{M}_3 = \mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}$. Then [65],

1. \mathbf{M}_3 has the distribution $W_b(n - a, \Sigma_3)$ and is independent of $(\mathbf{M}_{11}, \mathbf{M}_{22})$,
2. if $\Sigma_{12} = 0$, then $\mathbf{M}_{22} - \mathbf{M}_3 = \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}$ has the distribution $W_b(a, \Sigma_{22})$, and $\mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}$, \mathbf{M}_{11} , and \mathbf{M}_3 are jointly independent.

On the other hand, the hypothesis technique of the Union Intersection Test (UIT) uses the statistic based on the largest eigenvalue r_1^2 of \mathbf{M}_S .

However, \mathbf{M}_S can be written as $[\mathbf{M}_3 + (\mathbf{M}_{22} - \mathbf{M}_3)]^{-1}(\mathbf{M}_{22} - \mathbf{M}_3)$, where $\mathbf{M}_{22} = n\mathbf{S}_{YY}$, $\mathbf{M}_3 = n(\mathbf{S}_{YY} - \mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY})$, and $\mathbf{M}_{22} - \mathbf{M}_3$ satisfies the independence condition of the greatest root statistic. Therefore, under $H_0: \Sigma_{12} = 0$, r_1^2 has the distribution $\theta(p, n - q - 1, q)$, and the Tracy-Widom approximation can be applied.

The previous derivation shows a procedure for statistically determining the rank η of an RRR model through the RMT framework. Specifically, it has established the connection of the Tracy-Widom distribution to testing the null hypothesis $H_0: \Sigma_{12} = 0$ in the particular case of CCA for general RRR models. In what follows, the applied methodology for finding the number of significant components or factors in the CCA using our datasets of predictor and response cryptocurrency variables is described.

Analysis of predictor-response cryptocurrency factors

The first step in using these techniques with real-world data is based on numerically solving the system of equations in Eq (26), taking into account the Painlevé equations with the boundary condition that as $t \rightarrow \infty$, $q(t)$ asymptotically converges the Airy function $Ai(t)$. We solve these nonlinear differential equations with an relative tolerance error of 1×10^{-12} , following the approach described in [61]. Table 6 shows a subsample of the obtained values. The first and second columns display values of x, y on the plane of the Tracy-Widom distribution, respectively. The third column shows the cumulative density value (cdv) corresponding to the

Table 6. Tracy-Widom values.

x	y	cdv
⋮	⋮	⋮
1.995	0.017669	0.989510
2.000	0.017535	0.989598
2.005	0.017402	0.989685
2.010	0.017270	0.989771
2.015	0.017139	0.989857
2.020	0.017009	0.989942
2.025	0.016880	0.990026
2.030	0.016751	0.990110
2.035	0.016623	0.990193
2.040	0.016497	0.990276
⋮	⋮	⋮

Subsample of values x, y on a plane and the corresponding cdv of the Tracy-Widom distribution.

<https://doi.org/10.1371/journal.pone.0227269.t006>

respective values of x, y ; the result of subtracting that value from 1 determines the level of significance in the statistical test of Tracy-Widom.

Next, we apply CCA to the set of cryptocurrency variables. In this analysis, the predictor and response variables previously obtained in the section on variable selection are regarded as sets X, Y , respectively, where the response variables Y are led by a time step of $\Delta t = 1$ hour, as they were in the variable selection analysis. However, when applying the standard CCA, we observed numerical instabilities because some underlying variables were nearly collinear. To avoid this issue, the regularized version of the CCA is applied [66]. In this approach, covariance matrices S_{XX} and S_{YY} are replaced by

$$\begin{aligned} S_{XX} + \lambda I_x \\ S_{YY} + \lambda I_y, \end{aligned} \tag{33}$$

where I_x, I_y are identity matrices of the same dimensions as the corresponding covariance matrices S_{XX}, S_{YY} , and λ is a regularization parameter. Applying the regularized CCA ensures that the relevant covariance matrices have numerically stable inverse matrices. This transformation only affects the eigenvalues, which suffer a translation by the same amount λ , but the eigenvectors remain unchanged. Accordingly, we compute the eigenvalues of matrix M_s , considering a regularization parameter $\lambda = 0.01$. Then, under the greatest root distribution $\theta(p, n - q - 1, q)$ with parameters $p = 49$ (the number of predictor variables), $q = 51$ (the number of response variables), and $n = 4531$ (the number of observations after the time horizon reduction by considering lag-lead returns) using Eqs (28)–(31) results in obtaining 36 factors explaining 93.59% of variance at the significance level of $\alpha = 0.01$.

Fig 6 shows the percentage of explained variance as a function of the number of factors. Although, in CCA, the increments of the predictor and response components are considered symmetrically, the fixed lag time of $\Delta t = 1$ provides the predictive element. The dashed vertical gray line represents the cutoff point where the number of significant components is determined. The inset graph shows the same but on a semi-log scale. The plot does not show an abrupt change in the curve. Hence, if we use the elbow criterion, it would not be possible to determine the appropriate number of components to be considered in the model. Moreover,

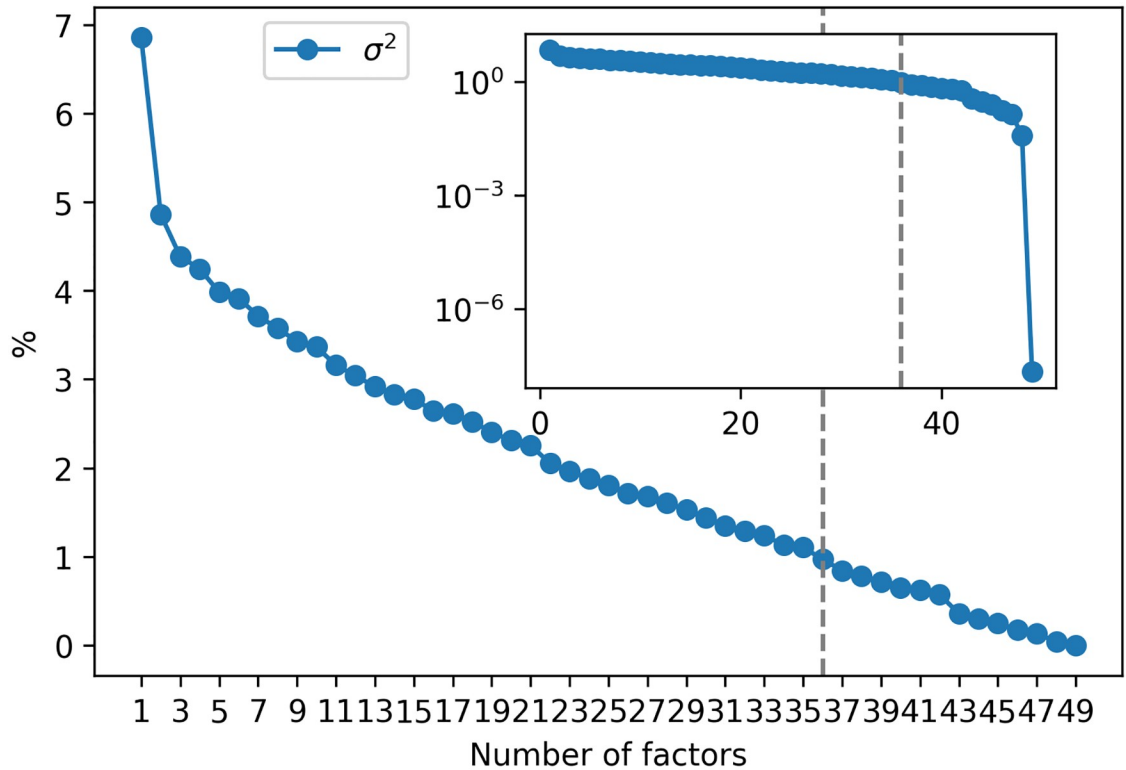


Fig 6. Explained variance in CCA. Percentage contribution to variance as a function of the component. The inset graph show the same but on a semi-log scale.

<https://doi.org/10.1371/journal.pone.0227269.g006>

compared with the resampling approach, the computational time of the Tracy-Widom test is negligible since we only need to compute the table of the significance level once.

In addition, Table 7 shows the number of factors and percentage of explained variance for $\lambda \in \{0.1, 0.01, 0.001, 1 \times 10^{-4}, 1 \times 10^{-5}\}$ and $\alpha \in \{0.1, 0.01, 0.001\}$. It is observed the number of factors oscillates between 34 and 37, while the corresponding explained variance is between 91.42% and 94.56%. Additionally, the number of factors seems to reach a stable value of 36 as λ, α become smaller.

Furthermore, the dynamic behavior of the number of factors can be analyzed. To this end, we partition the whole sample period and choose a set of subsample data matrices for four weeks ($n = 672$ hours) and a moving window of one week (168 hours). As a result, we obtain $m = 23$ subsample data matrices to which CCA is applied for the set of predictor-response variables obtained in the section on variable selection while considering the entire period of time.

Fig 7a shows the estimated number of factors as a function of the moving window for a fixed level of significance $\alpha = 0.01$ and various values of the regularization parameter $\lambda \in \{0.1, 0.01, 0.001, 1 \times 10^{-4}, 1 \times 10^{-5}, 1 \times 10^{-6}\}$, whereas Fig 7b presents the same analysis, but for a fixed value of $\lambda = 0.01$ and various levels of significance $\alpha \in \{0.1, 0.01, 0.001, 1 \times 10^{-4}, 1 \times 10^{-5}, 1 \times 10^{-6}\}$. In all cases, the estimation of the number of factors was performed using the same proposed procedure as in the analysis of the entire period (Eqs (28)–(31)). We observe in Fig 7a that the number of factors increases as the value of λ decreases, reaching a stable behavior at $\lambda = 1 \times 10^{-5}$, where the number of factors does not change for smaller values of λ and is thus similar to that observed at $\lambda = 1 \times 10^{-6}$. In contrast, in Fig 7b, the number of factors decreases as the value of α decreases, but the temporal behavior seems to have the same tendency as in

Table 7. Number of factors as a function of λ and α .

λ	α	Number of factors	Explained variance (%)
0.1	0.1	35	92.54
0.1	0.01	35	92.54
0.1	0.001	34	91.42
0.01	0.1	36	93.59
0.01	0.01	36	93.59
0.01	0.001	35	92.48
0.001	0.1	37	94.56
0.001	0.01	36	93.59
0.001	0.001	36	93.59
1×10^{-4}	0.1	37	94.56
1×10^{-4}	0.01	36	93.59
1×10^{-4}	0.001	36	93.59
1×10^{-5}	0.1	37	94.56
1×10^{-5}	0.01	36	93.59
1×10^{-5}	0.001	36	93.59

Number of factors and the respective explained variance (as a percentage) for various values of the regularization parameter λ and significance level α .

<https://doi.org/10.1371/journal.pone.0227269.t007>

Fig 7a. The fact that the number of factors decreases with the contraction of the time horizon is a natural behavior explained by Harding in [11]. Nonetheless, all cases exhibit a high value on November 24, 2018 independent of the magnitude.

To investigate a possible relation between the dynamics of the number of factors and the return of the exchange value of cryptocurrencies, two indices were constructed from the predictor and response variables. Let $I_x(l)$ be the composite index constructed by averaging over all the predictor variables during subsample periods $l = 1, \dots, m$. Similarly, we denote by $I_y(l)$ the composite index constructed in the same manner, but considering the response variables. **Table 8** shows the correlation values between these composite indices in relation with the

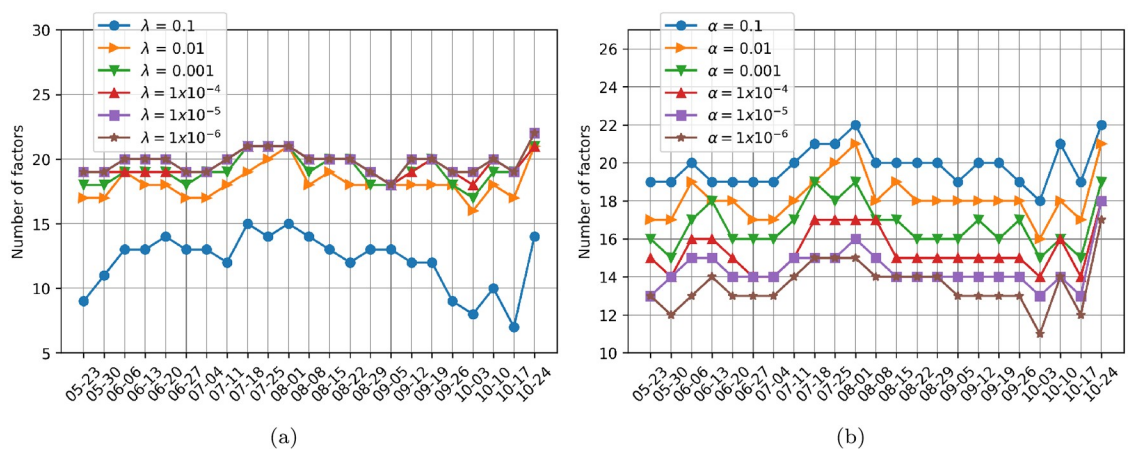


Fig 7. Number of factors as a function of time. Estimated factors for subsample data matrices over four weeks with a moving window of one week. (a) The level of significance is set at $\alpha = 0.01$, and the regularization parameter λ is varied. (b) The regularization parameter is set at $\lambda = 0.01$, and the significance level α is varied.

<https://doi.org/10.1371/journal.pone.0227269.g007>

Table 8. Correlations between I_x , I_y and the number of factors.

λ	α	Corr(I_x , #factors)	Corr(I_y , #factors)
0.1	0.01	-0.29	-0.27
0.01	0.01	-0.45	-0.44
0.001	0.01	-0.40	-0.41
1×10^{-4}	0.01	-0.43	-0.45
1×10^{-5}	0.01	-0.43	-0.44
1×10^{-6}	0.01	-0.43	-0.44
0.01	0.1	-0.42	-0.41
0.01	0.01	-0.45	-0.44
0.01	0.001	-0.46	-0.45
0.01	1×10^{-4}	-0.52	-0.51
0.01	1×10^{-5}	-0.49	-0.46
0.01	1×10^{-6}	-0.47	-0.45

Pearson correlations between the composite indices I_x , I_y and the number of factors (denoted by #factors). The same values of λ and α as in Fig 6 are considered.

<https://doi.org/10.1371/journal.pone.0227269.t008>

variation of the number of factors as a function of time. We observe a moderate anticorrelation in all cases, ranging from -0.29 to -0.52 . This result suggests that the variation in the number of factors over time is influenced by the behavior of the composite returns of the predictor and response cryptocurrencies. This result can be used in future work to modeling the collective behavior of the cryptocurrency market under a predictive model using as a predictor the time series created by the variation of the number of factors through time.

To identify the compositions of the predictor-response factors, i.e., if the weights contribute equally or if there are peaks where some cryptocurrencies have a dominant contribution, we compute the inverse participation ratio (IPR). This measure originates from physics, where it is used as a measure of localization [67]; however, in the context of econophysics, it allows us to determine the number of components that participate significantly in each eigenvector associated with the predictor-response factors. The IPR of eigenvector V_k is given by

$$IPR(k) = \sum_{j=1}^r |V_k(j)|^4, \tag{34}$$

where $k = 1, \dots, r$, and $r = p$ or q , depending on whether eigenvectors V_k are associated with the predictor or response factors, respectively. This quantity is bounded between $1/r$ and 1. If eigenvector V_k is localized only in one component, or equivalently, if only one coin has a dominant weight contribution, then $IPR(k) = 1$. In contrast, if V_k is not localized or is uniformly distributed over r weights, then $IPR_k = 1/r$. In Fig 8, we plot IPR for (a) the predictor factors and (b) the response factors of the entire sample period, varying the regularization parameter $\lambda \in \{0.1, 0.01, 0.001, 1 \times 10^{-4}, 1 \times 10^{-5}, 1 \times 10^{-6}\}$ and setting the level of significance to $\alpha = 0.01$. We do not change α in this analysis because doing so does not alter the results. We observe first that the behavior of IPR does not change significantly for either predictors or responses as λ is varied. A relevant point is that by considering IPR we can distinguish between the *localized factors*, i.e., those that have dominant weights related to specific cryptocurrencies, and the *nonlocalized factors* that have uniform coin contributions.

Thus, Fig 9 shows a plot of the weight contribution of the predictor and response factors with the highest IPR. Based on a visual inspection, we have chosen the predictor factors identified with $k = 2, 14, 40$ and 44 and the response factors identified with $k = 2, 49, 50$ and 51 as

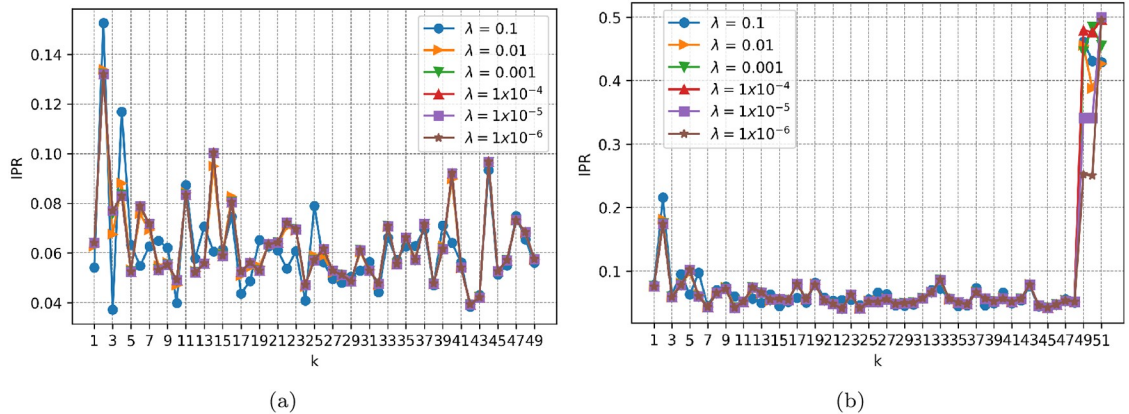


Fig 8. IPR. Inverse participation ratio for the entire sample period of time for various values of the regularization parameter λ . (a) Predictor factors. (b) Response factors.

<https://doi.org/10.1371/journal.pone.0227269.g008>

the more localized factors. As a result, returning to the preceding Fig 9, it is observed that there are a few cryptocurrencies with high weights, as expected, while the majority of cryptocurrencies oscillate around zero. This phenomenon is even more representative for the response factors of cryptocurrencies. For example, if we consider cryptocurrencies for which the absolute value of the weight is larger than 0.4, we observe that *tether*, *steem*, *bitshares*, *zilliqa* and *request network* (numbered 7, 8, 13, 27 and 44, respectively) are the more representative cryptocurrencies for the selected predictor factors, whereas cryptocurrencies *iota*, *zcash*, *raiblocks*, *maker*, *walton*, *iostoken* and *gxshares* (numbered 8, 16, 17, 24, 25, 41 and 42, respectively) are the equivalent for the response factors (see S1 Table for reference). In general, these factors fluctuate around zero but have a strong peak for a specific currency. These are the corresponding relevant currencies in a predictive model and has to do with the more and less diversified portfolio, and its association with the risk of investing in a portfolio. Specifically, we could explore investing in a mixed portfolio for a predictive model. It can be said that they are the components with the greatest weight that are involved in the predictive model. A model could be tested only with those components as future work.

Additionally, Fig 10 shows a plot of the dynamic behavior of IPR for each $k = 1, \dots, r$, denoted by IPR_k , as a function of time using the same set of subsample data matrices

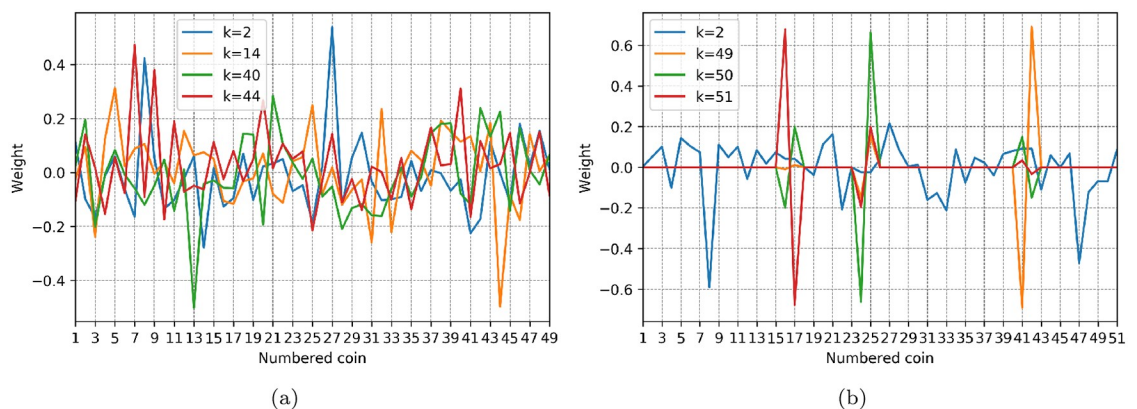


Fig 9. Factor weights. Contribution weight of each coin for the factors with the highest IPR. (a) Eigenvector weights associated with the predictor factors $k = 2, 14, 40$ and 44 . (b) Eigenvector weights associated with the response factors $k = 2, 49, 50$ and 51 .

<https://doi.org/10.1371/journal.pone.0227269.g009>

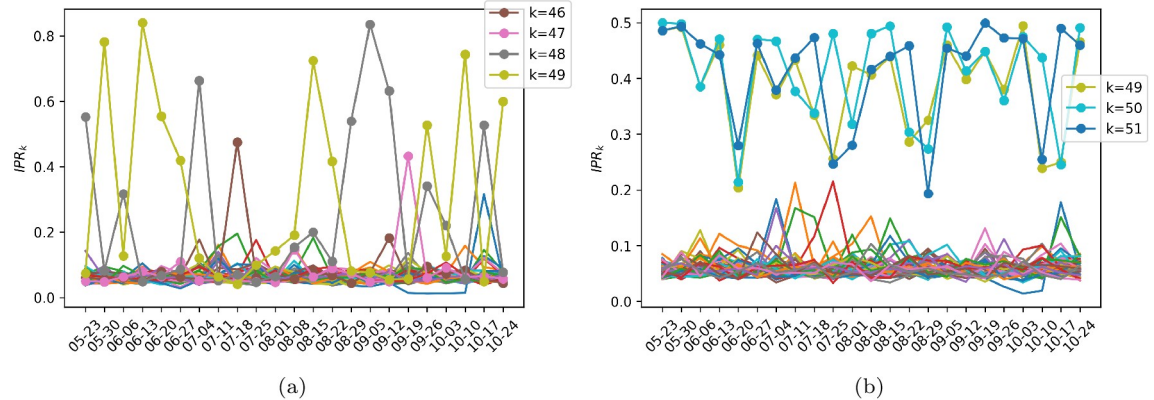


Fig 10. IPR as a function of time. Dynamic inverse participation ratio for the set of subsample data matrices described above. The regularization parameter and level of significance are set to $\lambda = 0.01$ and $\alpha = 0.01$, respectively. (a) Predictor factors. (b) Response factors.

<https://doi.org/10.1371/journal.pone.0227269.g010>

constructed above, where again $r = p$ for the predictor factors and $r = q$ for the response factors, and settings $\alpha = 0.01$ and $\lambda = 0.01$ are used. It is observed in general that the last factors are more *localized* in the sense of IPR. For example, the last predictor factors 46, 47, 48 and 49 show the highest IPR values consistently over time (see Fig 10a), while the last response factors 49, 50 and 51 exhibit this behavior as well (see Fig 10b). Based on these results, we can call these last factors the *specific-predictor* and *specific-response* modes, respectively.

To finalize this analysis, it is important to verify if the factors are correlated in the estimated sample because highly correlated predictor factors provide redundant information about the response and can lead to overfitting. Therefore, Fig 11a and 11b show, for the entire sample period, the normalized distribution $\rho(\sigma_{ij})$ of correlations between all pairs of factors σ_{ij} , $i, j = 1, \dots, r$, where $i \neq j$, and $r = p$ for the predictor factors, and $r = q$ for the response factors, respectively. Similarly, Fig 11c and 11d show, for the set of subsample data matrices, the normalized distribution $\rho(\sigma_{ij}^m)$ of correlations between all pairs of factors σ_{ij}^m , $i, j = 1, \dots, r$ and $m = 1, \dots, 23$, where $i \neq j$, and again, $r = p$ for the predictor factors, and $r = q$ for the response factors, respectively. In this last case, symbol m represents that the normalized distributions $\rho(\sigma^m)$ of correlations σ^m have taken into account $m = 23$ subsample matrices (constructed using the entire six-month sample) to calculate the densities. In all shown cases, settings $\alpha = 0.01$ and $\lambda = 0.01$ have been used. In general, Fig 11 does not show significant correlations in any case. Hence, the estimated predictor and response factors are correlated in neither the estimation sample nor the considered subsample windows. Therefore, the predictor factors do not provide redundant information about the response in the predictive model, which out-of-sample analysis does not exploit here because it falls outside the scope of this study and is left for a future study, where different Γ structures will be considered (see Eq 21).

Conclusions

In general, random matrices seem to be a promising tool for performing factor determination in financial and economic problems. Nevertheless, extensive theoretical studies related to random matrices have not yet been applied by practitioners. In part, this lack of use is due to the absence of a simple statistical test that provides reliable results. Accordingly, to fill this gap, we describe the connection between the RRR models and the Tracy-Widom test used to determine the number of significant factors or components in the reduced CCA case of the general

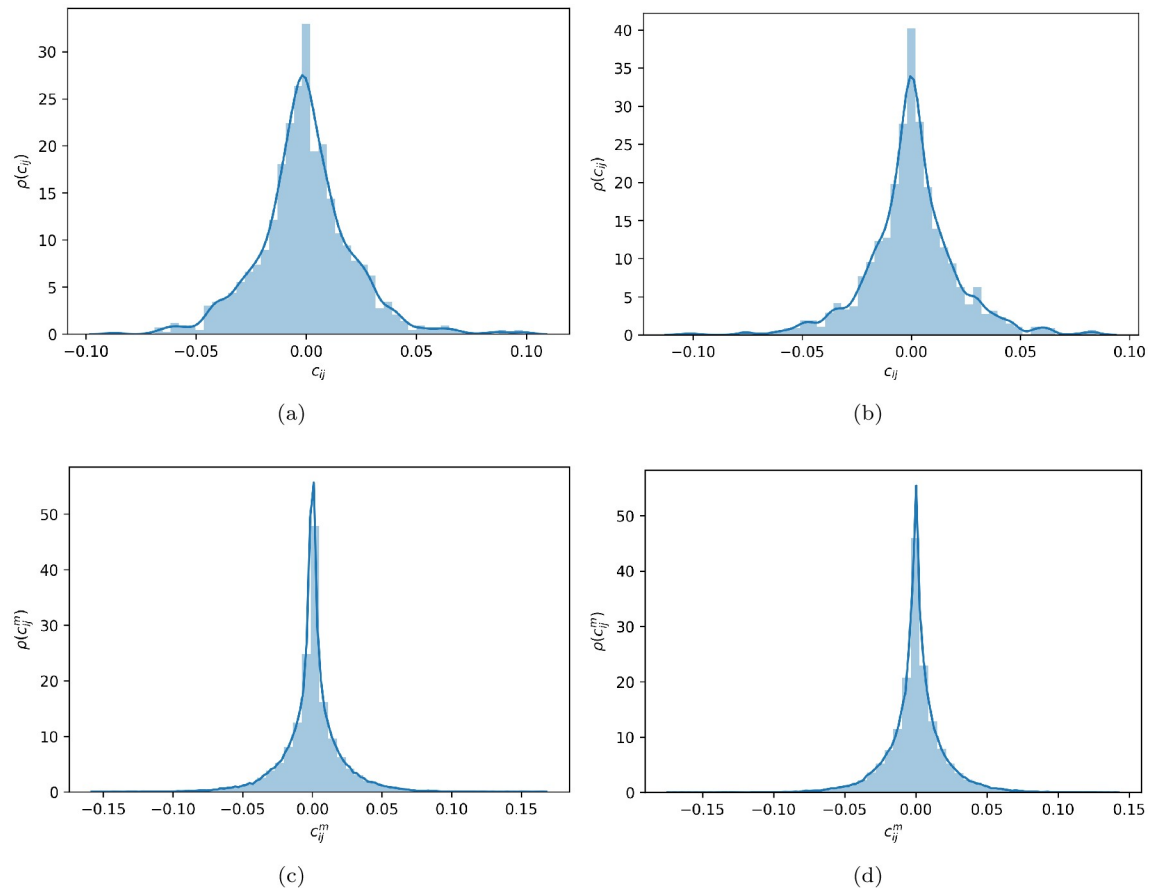


Fig 11. Distribution of correlations between factors. (a) Correlations between predictor factors determined by considering the entire sample period. (b) Correlations between response factors determined by considering the entire sample period. (c) Correlations between predictor factors determined by considering all the subsample components. (d) Correlations between response factors determined by considering all the subsample components. In all cases, settings $\alpha = 0.01$ and $\lambda = 0.01$ have been used.

<https://doi.org/10.1371/journal.pone.0227269.g011>

RRR models. The main advantage of the proposed procedure is avoiding the subjective element of visual inspection and the computational cost of the resampling approach. Furthermore, the distributional Tracy-Widom test is conceptually related to a more general mathematical framework and involves many branches of basic mathematics and theoretical physics.

In particular, the results obtained here show that a large number of factors are statistically significant in the dataset of cryptocurrencies. Hence, contrary to what a visual inspection suggests, this methodology allows for the retention of more factors. Moreover, the dynamic behavior of the number of factors seems to be related with the dynamic of the composite index of predictor and response cryptocurrencies. As such, future research in this direction could yield insights allowing one to anticipate drops in exchange rates. Furthermore, an IPR analysis suggests the existence of specific-predictor and specific-response factors. These factors are the eigenvectors associated with the smallest singular values obtained using CCA, where only a few cryptocurrencies have remarkable weights and account for the dynamic.

Another contribution of this study is the variable selection methodology based on information theory. We used TE to measure the flow of information between cryptocurrencies' return

variables and specify the predictor and response variables. Since TE can be regarded as a generalization of the Granger causality test under some circumstances, we can cover many scenarios, including possible nonlinear dependencies between the variables. Actually, the STE estimation shows better performance over Granger causality test for simulated systems with linear and nonlinear interactions. The STE can detect the cluster structure, while Granger causality fails to discriminate between the modeled interaction blocks and indicate a higher number of spurious causal dependencies.

In the application to cryptocurrencies we propose a heuristic criterion related to the in-degree and out-degree of the nodes observed if the TE estimation is treated as a graph. Again, the symbolic approach to estimating TE has the advantage of having a distributional test. Therefore, our selected set of response and predictor variables has an associated *p-value*, which is always preferred in the econometric community and makes our results more robust from the statistical perspective. In addition, implementation through the symbolic version is less expensive computationally compared to other approaches such as binning and kernels, and consequently can be used in a large dimensional data set.

The modularity and centrality results give us an understanding of the structure of cryptocurrencies that do not emerge at first glance. We have confirmed that our selected response and predictor variables have a natural cluster structure at least under the modularity approach. Additionally, we have found that modularity validates our selection rule of the sets of predictive and predicted variables. It has also been found that cryptocurrencies called tokens are central in the sense of sending information, while coins are central in the sense of receiving information, in addition to being the authorities under the HITS algorithm. This gives us glimpses of new structures that had not been reported before in the literature and provides a better understanding of the characteristics of cryptocurrencies beyond their predictive power.

An interesting future research direction is to explore other RRR models with a different Γ structure and explore the consequences of their predictive power in different scenarios. Similarly, it is exciting to explore further the clustering structure of the cryptocurrencies under the multivariate STE. In addition, its consequences in the theory of portfolios using the significant predicted and predictor eigenvalues to minimize the risk associated with an investment scenario. Finally, it is important to study the case of $\Sigma \neq \mathbf{I}$ to take into account spatio-temporal correlations. The last problem is related to the well-known dynamic factor models in the econometric literature and has the advantage of being more explanatory and linked with structural predictive models such as vector autoregressive (VAR) and vector error correction models (VECM).

Supporting information

S1 File. Raw data. Prices of cryptocurrencies used in this study as described in the Data section before preprocessing.
(XLSX)

S1 Table. List of cryptocurrency names. Cryptocurrencies are listed in the descending order of capitalization as of February 2018.
(PDF)

S2 Table. Entire list of predictor-response variables. Each set is ordered from the highest to the lowest capitalization of included cryptocurrencies.
(PDF)

Acknowledgments

The authors would like to thank Zhiya Zuo for providing the modularity implementation, and Antonio Álvarez for his invaluable experience in data acquisition. Also, we need to thanks the reviewers for their insights and comments that helped improve the quality of the paper.

Author Contributions

Conceptualization: Andrés García-Medina.

Data curation: Andrés García-Medina.

Formal analysis: Andrés García-Medina.

Funding acquisition: Graciela González Farías.

Investigation: Andrés García-Medina.

Methodology: Andrés García-Medina.

Project administration: Graciela González Farías.

Validation: Andrés García-Medina.

Visualization: Andrés García-Medina.

Writing – original draft: Andrés García-Medina.

Writing – review & editing: Andrés García-Medina, Graciela González Farías.

References

1. Cattell RB. The Scree Test For The Number Of Factors. *Multivar. Behav. Res.* 1966; 1(2):245–276. https://doi.org/10.1207/s15327906mbr0102_10
2. Wold S. Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics.* 1978; 20(4):397–405. <https://doi.org/10.1080/00401706.1978.10489693>
3. Zientek LR, Thompson B. Applying the bootstrap to the multivariate case: Bootstrap component/factor analysis. *Behav Res Methods.* 2007; 39(2):318–325. <https://doi.org/10.3758/bf03193163> PMID: 17695360
4. Kaiser HF. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* 1960; 20: 141–151. <https://doi.org/10.1177/001316446002000116>
5. Guttman L. Some necessary conditions for common factor analysis. *Educ. Psychol. Meas.* 1960; 20: 141–151.
6. Braeken J, Van Assen MA. An empirical Kaiser criterion. *Psychological Methods.* 2017; 22(3):450. <https://doi.org/10.1037/met0000074> PMID: 27031883
7. Kapetanios G. A new method for determining the number of factors in factor models with large datasets. Queen Mary University of London, School of Economics and Finance. 2004; Working Paper 525.
8. Onatski A. Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* 2010; 92(4):1004–1016. https://doi.org/10.1162/REST_a_00043
9. Lam C, Yao Q. Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Stat.* 2012; 40(2):694–726. <https://doi.org/10.1214/12-AOS970>
10. Tracy CA, Widom H. Spacing distributions and the Airy kernel. *Comm. Math. Phys.* 1994 Jan 1; 159(1):151–74. <https://doi.org/10.1007/BF02100489>
11. Harding M. Estimating the number of factors in large dimensional factor models. *J Econom.* 2013.
12. Yeo J, Papanicolaou G. Random matrix approach to estimation of high-dimensional factor models; 2016. Preprint. Available from: arxiv.org/abs/1611.05571. Cited 10 July 2019.
13. Johnstone IM. Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Stat.* 2008; 36(6):2638. <https://doi.org/10.1214/08-AOS605> PMID: 20157626
14. Mantegna RN., Stanley HE. Scaling behaviour in the dynamics of an economic index. *Nature.* 1995; 376(6535):46. <https://doi.org/10.1038/376046a0>

15. Burda Z, Jurkiewicz J, Nowak MA, Papp G, Zahed I. Free random Lévy variables and financial probabilities. *Physica A*. 2001; 299(1-2):181–187. [https://doi.org/10.1016/S0378-4371\(01\)00294-1](https://doi.org/10.1016/S0378-4371(01)00294-1)
16. Burda Z, Jurkiewicz J, Nowak MA, Papp G, Zahed I. Free Lévy matrices and financial correlations. *Physica A*. 2004; 343:694–700. <https://doi.org/10.1016/j.physa.2004.05.049>
17. Biroli G, Bouchaud JP, Potters M On the top eigenvalue of heavy-tailed random matrices. *Europhysics Letters*. 2007; 78(1):10001. <https://doi.org/10.1209/0295-5075/78/10001>
18. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. Bitcoin. 2009. Available from <https://bitcoin.org/bitcoin.pdf> Cited 26 April 2019.
19. Cryptocurrency Market Capitalizations. CoinMarketCap. Available from <https://coinmarketcap.com> Cited 26 April 2019. It has been complied with the terms of service of the CoinMarketCap API.
20. Fama EF. The behavior of stock-market prices. *J. Bus.* 1965; 38(1):34–105. <https://doi.org/10.1086/294743>
21. Urquhart A. The inefficiency of Bitcoin. *Econ. Lett.* 2016; 148:80–82. <https://doi.org/10.1016/j.econlet.2016.09.019>
22. Alessandretti L, ElBahrawy A, Aiello LM, Baronchelli A. Anticipating cryptocurrency prices using machine learning. *Complexity*. 2018. <https://doi.org/10.1155/2018/8983590>
23. Caporale GM, Gil-Alana L, Plastun A. Persistence in the cryptocurrency market. *Res. Int. Bus. Finance*. 2018; 46: 141–148. <https://doi.org/10.1016/j.ribaf.2018.01.002>
24. Stosic D, Stosic D, Ludermir TB, Stosic T. Collective behavior of cryptocurrency price changes. *Physica A*. 2018 Oct 1; 507:499–509. <https://doi.org/10.1016/j.physa.2018.05.050>
25. Marchenko VA, Pastur LA. Distribution of eigenvalues for some sets of random matrices. *Sb. Math.* 1967; 114(4):507–36.
26. Begušić S, Kostanjčar Z, Stanley HE, Podobnik B. Scaling properties of extreme price fluctuations in Bitcoin markets. *Physica A*. 2018 Nov 15; 510:400–6. <https://doi.org/10.1016/j.physa.2018.06.131>
27. Schreiber T. Measuring information transfer *Phys. Rev. Lett.* 2000; 85(2):461. <https://doi.org/10.1103/PhysRevLett.85.461> PMID: 10991308
28. Shannon CE. A mathematical theory of communication *Bell Syst. Tech. J.* 1948; 27(3):379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
29. Prokopenko M, Lizier JT. Transfer entropy and transient limits of computation *Sci. Rep.* 2014; 4:5394.
30. Papan A, Kugiumtzis D, Larsson PG. Reducing the bias of causality measures *Phys. Rev. E*. 2011; 83(3):036207. <https://doi.org/10.1103/PhysRevE.83.036207>
31. Barnett L, Bossomaier T. Transfer entropy as a log-likelihood ratio. *Phys. Rev. Lett.* 2012 Sep 28; 109(13):138105. <https://doi.org/10.1103/PhysRevLett.109.138105> PMID: 23030125
32. Liang X. The Liang-Kleeman information flow: Theory and applications. *Entropy*. 2013; 15(1):327–360. <https://doi.org/10.3390/e15010327>
33. Prokopenko M, Lizier J, Price D. On thermodynamic interpretation of transfer entropy. *Entropy*. 2013; 15(2):524–543. <https://doi.org/10.3390/e15020524>
34. Barnett L, Barrett AB, Seth AK. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* 2009 Dec 4; 103(23):238701. <https://doi.org/10.1103/PhysRevLett.103.238701> PMID: 20366183
35. Marschinski R, Kantz H. Analysing the information flow between financial time series *Eur. Phys. J. B*. 2002; 30(2): 275–281. <https://doi.org/10.1140/epjb/e2002-00379-2>
36. Sandoval L. Structure of a Global Network of financial companies based on transfer entropy. *Entropy*. 2014; 16(8):4443–4482. <https://doi.org/10.3390/e16084443>
37. Begušić S, Kostanjčar Z, Kovač D, Stanley HE, Podobnik B. Information Feedback in Temporal Networks as a Predictor of Market Crashes. *Complexity*. 2018; 1–13.
38. Staniak M, Lehnertz K. Symbolic Transfer Entropy. *Phys. Rev. Lett.* 2008 Apr 14; 100(15):158101. <https://doi.org/10.1103/PhysRevLett.100.158101> PMID: 18518155
39. Papan A, Kyrtsov C, Kugiumtzis D, Diks C. Detecting causality in non-stationary time series using partial symbolic transfer entropy: evidence in financial data. *Comput. Econ.* 2016; 47(3):341–365. <https://doi.org/10.1007/s10614-015-9491-x>
40. Dickey DA, Fuller WA. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *J. Am. Stat. Assoc.* 1979 Jun 1; 74(366a):427–31. <https://doi.org/10.2307/2286348>
41. Kullback S. *Information Theory and Statistics*. 1st ed. New York: Joh Wiley & Sons; 1959.
42. Granger CW. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. 1969 Aug 1:424–38. <https://doi.org/10.2307/1912791>

43. Bradley E, Kantz H. Nonlinear time-series analysis revisited *Chaos* (2015); 25(9):097610. <https://doi.org/10.1063/1.4917289> PMID: 26428563
44. Bossomaier T, Barnett L, Harré M, Lizier JT. *An Introduction to Transfer Entropy: Information Flow in Complex Systems*. Cham, Germany: Springer International Publishing; 2016.
45. Hlaváčková-Schindler K. Equivalence of Granger causality and transfer entropy: A generalization. *Appl. Math. Sci.* 2011; 5(73):3637–48.
46. Bandt C, Keller G, Pompe B. Permutation Entropy: A Natural Complexity Measure for Time Series Non-linearity. 2002 Apr 11; 88(17):174102.
47. Lizier JT. JIDT: An Information-Theoretic Toolkit for Studying the Dynamics of Complex Systems. *Front. robot. AI.* 2014 Dec 2; 1:11. <https://doi.org/10.3389/frobt.2014.00011>
48. Newman MEJ. CModularity and community structure in networks. *PNAS.* 2006; 103(23):8577–82. <https://doi.org/10.1073/pnas.0601602103> PMID: 16723398
49. Leicht EA, Newman MEJ. Community Structure in Directed Networks. *J.Phys. Rev. Lett.* 2008 Mar 21; 100(11):118703. <https://doi.org/10.1103/PhysRevLett.100.118703>
50. Fruchterman TMJ, Reingold EM Graph Drawing by Force-directed Placement. *Software-practice and experience.* 1991 Nov, 21(11):1129–1164. <https://doi.org/10.1002/spe.4380211102>
51. Newman MEJ *Networks, and introduction*. New York, NY, USA: Oxford University Press; 2010.
52. Kleinberg JM. Authoritative sources in a hyperlinked environment. *J. ACM* 1999; 46(5):604–632. <https://doi.org/10.1145/324133.324140>
53. Izenman AJ. Reduced-rank regression for the multivariate linear model. *J. Multivar. Anal.* 1975 Jun 1; 5(2):248–64. [https://doi.org/10.1016/0047-259X\(75\)90042-1](https://doi.org/10.1016/0047-259X(75)90042-1)
54. Izenman AJ. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning* 1st ed. New York: Springer-Verlag; 2008.
55. Guhr T, Müller-Groeling A, Weidenmüller HA. Random-matrix theories in quantum physics: common concepts. *Phys. Rep.* 1998 Jun 1; 299(4-6):189–425. [https://doi.org/10.1016/S0370-1573\(97\)00088-4](https://doi.org/10.1016/S0370-1573(97)00088-4)
56. Forrester PJ, Huges TD. Complex Wishart matrices and conductance in mesoscopic systems: Exact results. *J. Math. Phys.* 1994 Dec; 35(12):6736–47. <https://doi.org/10.1063/1.530639>
57. Lou F, Zhong J, Yang Y, Zhou J. Application of random matrix theory to microarray data for discovering functional gene modules. *Phys. Rev. E.* 2006 Mar 29; 73(3):031924. <https://doi.org/10.1103/PhysRevE.73.031924>
58. Telatar E. Capacity of Multi-antenna Gaussian Channels. *Eur.Trans.Telecomm.* 1999 Nov; 10(6):585–95. <https://doi.org/10.1002/ett.4460100604>
59. Plerou V, Gopikrishnan P, Rosenow B, Amaral LA, Stanley HE. Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series. *Phys. Rev. Lett.* 1999 Aug 16; 83(7):1471. <https://doi.org/10.1103/PhysRevLett.83.1471>
60. Laloux L, Cizeau P, Bouchaud JP, Potters M. Noise Dressing of Financial Correlation Matrices. *Phys. Rev. Lett.* 1999 Aug 16; 83(7):1467. <https://doi.org/10.1103/PhysRevLett.83.1467>
61. Edelman A, Wang Y. *Random Matrix Theory and Its Innovative Applications*. In: Melnik R, Kotsireas IS, editors. *Advances in Applied Mathematics, Modeling, and Computational Science*. Boston: Springer; 2013. pp. 91–116.
62. Johnstone IM. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Stat.* 2001; 29(2):295–327. <https://doi.org/10.1214/aos/1009210544>
63. Johnstone IM. High dimensional statistical inference and random matrices; 2006. Preprint. Available from: arXiv:math/0611589. Cited 26 April 2019.
64. Johnstone IM. Approximate null distribution of the largest root in multivariate analysis. *Ann. Appl. Stat.* 2009; 3(4):1616. <https://doi.org/10.1214/08-AOAS220> PMID: 20526465
65. Mardia KV, Kent JT, Bibby JM. *Multivariate Analysis*. 1st ed. London: Academic Press; 1979.
66. Vinod HD. Canonical ridge and econometrics of joint production. *J. Econometrics* 1976; 4(2):147–166. [https://doi.org/10.1016/0304-4076\(76\)90010-5](https://doi.org/10.1016/0304-4076(76)90010-5)
67. Visscher WM. Localization of electron wave functions in disordered systems *J. Non-Cryst. Solids* (1972);(8): 477–484. [https://doi.org/10.1016/0022-3093\(72\)90179-2](https://doi.org/10.1016/0022-3093(72)90179-2)