Physica A 537 (2020) 122574

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Causal inference between cryptocurrency narratives and prices: Evidence from a complex dynamic ecosystem

Andrés Azqueta-Gavaldón[1]

University of Glasgow, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Article history:
Available online 19 September 2019

JEL classification:
C4
G1
C11

Keywords:
Narratives
Cryptocurrencies
Sentiment analysis
Latent dirichlet allocation (LDA)
Complex dynamic systems
Convergent cross mapping (CCM)

ABSTRACT

In this note, I explore the causal relationship between narratives propagated by the media and crypto prices. Firstly, I unveil four cryptocurrency-related narratives: *investment*, *technological innovation*, *security breaches* and *regulation*. Secondly, after acknowledging their tone (sentiment), I apply Convergent Cross Mapping (CCM) to assess the causal relationship between narratives and prices. I find strong bi-directional causal relationships between narratives concerning investment and regulation while a uni-directional causal association exists in narratives relating technology and security to prices. Therefore, this work contributes to the recent economic literature that connects consumer behaviour to narratives.

1. Introduction

> "*The best examples now of irrational exuberance or speculative bubbles is bitcoin. And I think that has to do with the motivating quality of the bitcoin story.*" Robert Shiller, 05 Sept 2017.[2]

There is a growing acknowledgement that narratives have an impact on economic activity [2,3]. "*Stories motivate and connect activities to deeply felt values and needs. Narratives 'go viral' and spread far, even worldwide, with economic impact*" [3]. It is therefore debatable as to what extent narratives are responsible for the recent exceptional volatility in cryptocurrency prices. For example, Bitcoin prices went from $2000 in July 2017 to almost $20,000 by December of the same year before falling to $6000 in April 2018. While some individuals see Bitcoin as a fad and an example of irrational exuberance or speculative bubble [1], a much more enthusiastic view also co-exists; that Bitcoin represents fundamental transformation of money where transactions have no state [4].[3] In addition, technological innovations behind cryptocurrencies (such as the blockchain) have also generated excitement.

Establishing to what extent this narrative diversity has caused price changes is the main endeavour of this paper. The relationship between narratives and prices ought to be driven by complex interactions. For example, articles written in the

E-mail address: a.azqueta-gavaldon.1@research.gla.ac.uk.

1 I thank George Sugihara, Charles Nolan, Campbell Leith, Pablo Herrero, and four anonymous referees for their comments and suggestions. I also thank Elisa Castagno for comments and proofreading this work. I, alone, am responsible for any errors.

2 [1].

3 "*Bitcoin represents a fundamental transformation of money where a transaction has no state or context other than obeying the consensus rules of the network that no-one controls. Where your money is yours. It is a system of money that is simultaneously, absolutely transnational and borderless*". Antonopoulos [4].

https://doi.org/10.1016/j.physa.2019.122574

media about a specific phenomenon will attract or detract new investors depending on their content and tone (sentiment). In this case, the media might be a driver of prices. However, the press might also play a booster role; it reacts to price changes by increasing the coverage of a given topic. For this reason, one phenomenon cannot be understood without the other, indicating the *non-separability* of the system. As such, one of the main challenges that arises when examining non-separable ecosystems is how to find the tools that are best suited to their study.

To formally test the relationship between narratives and prices, I make use of a relatively new causal inference method suited to complex dynamical systems: Convergence Cross Mapping (CCM). CCM relies on state-space reconstructions meaning that it directly recovers the dynamics of the system from the time series data, without assuming any set of equations governing the system. CCM infers patterns and associations from the data instead of using a set of parametric equations that might be impractical when the exact mechanisms are unknown or too complex to be characterised with existing datasets. The intuition behind these kind of more flexible models is that if the dynamics of one variable can be forecasted by the time print of the other, there is a causal relationship. Each variable can identify the state of the other in the same way that information about past prey populations can be recovered from the predator time series, and vice versa [5]. Unlike other convectional techniques like the Granger causality test, CCM does not assume a pure stochastic system where variables are totally independent (separable) from each other.

To quantify the propagation of the main narratives, I use an unsupervised machine-learning algorithm on news-media articles that contain words related to cryptocurrencies. The algorithm is unsupervised in the sense that it infers the themes of a set of documents without any need for labelling the articles or training the model before the articles are classified. With the help of this algorithm, I unveil four distinctive narratives running from April 2013 to December 2018. These narratives describe events related to *investment*, *technology*, *crime*, and *regulation*. Unsurprisingly, the first two of these display a more positive sentiment than do the latter two. Additionally, while the first two narratives rise during sharp increases in cryptocurrency prices, it is noted that the latter two do so during price stagnation.

Overall, I find strong bi-directional causal relationships between narratives and cryptocurrency prices. That is, price dynamics influence the propagation of news articles describing the cryptocurrency phenomenon while, simultaneously, narratives influence price dynamics. However, the strength of these causal relationships is not homogeneous among the various narratives. Results suggest that cryptocurrency prices have the strongest causal impact on news relating to investment and regulation and the least impact on news relating to technology or security issues. The former phenomenon can be explained by the fact that price changes directly affect investment (either positively or negatively) while putting pressure on policymakers to adopt new regulations. I also find that the investment narratives affect price dynamics, although the strength of the relationship is lower than that from prices causing narratives. Therefore, the press seems to act as a signal booster for events relating to investments as it reacts to price dynamics by describing the investment side, leading to further auspicious changes in prices. A similar situation occurs with the Regulation narrative; this is also found to influence prices, albeit at a lower degree than prices influencing the regulatory narrative.

These results hold up to a battery of robustness tests. Firstly, I use the randomisation tests together with a surrogate time series to test for significance in these causal relationships [6]. This test compares the output generated by the CCM for the actual model and an alternative one generated trough surrogate time series. Secondly, I account for time-delayed causal interactions in order to assess whether this ecosystem suffers from "generalised synchrony". Generalised synchrony occurs when there is an exceptionally strong unidirectional force in the ecosystem that makes the response variable to collapse in the driving variable, therefore producing a false sense of causal interaction. However, this is not the case in our ecosystem, since the results hold up when I apply a monthly lag to either the narratives or the prices. Finally, I compare the CCM results to alternative causality approaches such as the Granger causality test.

This paper relates to at least two strands of literature. The first is research into the relationship between news and asset prices. Go and Ederington [7] have already documented that negative news associated with deteriorating financial prospects has an effect on stock returns. Regarding the effect of positive versus negative news, we find two contradictory results. On the one hand, Bomfim [8] has found that positive surprises affecting the monetary policy target (news) tend to have a larger effect on volatility than do negative surprises. On the other hand, Gande and Parsley [9] found that, while sovereign spreads did not react to positive news (positive ratings), they did react to negative ones.

Second, research exists that studies the price dynamics of cryptocurrencies. Empirical work on this topic dates to Kristoufek [10], who found a strong link between queries on Google Trends or Wikipedia and Bitcoin prices. Additionally, Garcia et al. [11] have identified two positive feedback loops that led to Bitcoin price bubbles: one driven by word of mouth and the other by new Bitcoin adopters. Yelowitz and Wilson [12] have collected Google Trends data to reveal four possible Bitcoin user profiles: computer programming enthusiasts, speculative investors, libertarians and criminals. Phillips and Gorse [13] document the relationship between Bitcoin price changes and topical discussions on social media. Lastly, Begušić et al. [14] point out that Bitcoin returns, in addition to being more volatile, also exhibit heavier tails than do stocks.

The rest of the paper proceeds as follows: Section 2 describes the algorithm and data used to uncover the narratives relating to cryptocurrencies. Section 3 presents the empirical framework used to study the causal effects of narratives and prices and gives a description of the data. Section 4 shows the empirical findings, while Section 5 concludes.

Table 1
Cryptocurrency narratives.

Category	*Label*	*%*	*S*	*Top words LDA*
Financial Investment	Investment I	16.4	0.07	*cryptocurr, fund, trade, investor, market, sec, invest, exchang, offer, coin, etf, token, bitcoin, crypto, futur, asset, firm, ico, accord, ethereum*
	Investment II	12.8	0.07	*stock, bond, market, year, rate, china, quarter, billion, growth, oil, expect, rose, price, index, yield, analyst, investor, profit, sale, gain*
Technology	Technology	12.5	0.09	*blockchain, technolog, compani, sturtup, ventur, busi, partner, use, inc, build, nvidia, fintech, work, firm, develop, ledger, ebay, tech, million, invest, silicon*
Regulation	Polit. Regu.	12.5	0.07	*trump, presid, state, elect, polit, payment, govern, democrat, republican, rule, administr, senat, polici, vote, countri, american, washington, obama, congress*
	Financ. Regu.	12.3	0.06	*bitcoin, currenc, virtual, bank, exchang, payment, regul, money, transact, servic, central, financi, launder, account, withdraw, trade, deposit, rule, merchant, tax*
Security	Security	7.4	0.05	*attack, india, hacker, ransomwar, comput, hack, indian, cyber, secur, data, breach, cybersecur, ransom, north, victim, softwar, wannacri, rs, system, target, infect*
	Crime	5.5	0.03	*mt, gox, silk, ulbricht, prosecutor, road, shrem, arrest, karpel, indict, allegedli, charg, court, bankrupci, crimin, enforc, liberti, tokyo, lawyer, complaint, bitcoin*
				Other topics
	Assets	12	0.09	*bitcoin, gold, valu, like, bubbl, money, mine, miner, even, say, currenc, time*
	Corporations	6.2	0.1	*phone, facebook, googl, wilson, video, tesla, musk, app, twitter, word, privaci*
	Unknown	2.5	0.09	*craig, dimond, jewelri, christi, die, gem, student, incorrectli, collector, art, ira*

Notes: Most representative words for each topic display by the LDA algorithm, labelling of each narrative, Sentiment score (S), and percentage of each narrative on the corpus (%). Individual narratives are grouped into four broader categories.

2. Media coverage narratives

To obtain the narratives related to cryptocurrencies, I use a machine learning algorithm called *Latent Dirichlet Allocation* (LDA). LDA was developed by Blei et al. [15] and is a generative probabilistic model that labels textual documents with a distribution of topics while each topic consists of a distribution of words. The model recovers these two distributions by obtaining those model parameters that maximise the probability of each word appearing in each article given the total number of topics K. The probability of word w_i occurring in an article is:

$$P(w_i) = \sum_{j=1}^{K} P(w_i|z_i = j)P(z_i = j) \tag{1}$$

where z_i is a latent variable indicating the topic from which the ith word was drawn and $P(w_i|z_i = j)$ is the probability of word w_i being drawn from topic j. $P(z_i = j)$ is the probability of drawing a word from topic j in the current article, which will vary across different articles. The goal is, therefore, to maximise $P(w_i|z_i = j)$ and $P(z_i = j)$ from Eq. (1). However, direct maximisation turns out to be susceptible to finding local maxima and slow convergence [16]. To overcome this issue, I use online variational Bayes as proposed by Hoffman et al. [17]. This method approximates the posterior distribution of $P(w_i|z_i = j)$ and $P(z_i = j)$ using an alternative and simpler distribution: $P(z|w)$, and associated parameters.[4]

I run the LDA algorithm for all news articles that describe cryptocurrencies (those containing any form of the terms *bitcoin* or *cryptocurrency*) from worldwide business press: *The Financial Times*, *The Economist*, *The Economic Times*, *Business Insider* and *The Wall Street Journal*. The total number of news articles associated with any form of these two terms from March 2013 to December 2018 was 4503. Consistent with previous studies, I filter the textual data by removing stopwords (e.g. me, or, the, a) and uni-characters, convert all words into lower cases, and transform each word into its root (stemming). In this way, I reveal ten topics in this corpus.[5]

Table 1 shows all the 10 narratives revealed by the LDA. The two largest of all correspond to narratives related to financial investment, in which we find words such as *trade*, *investor*, *market*, *asset* or *ico* (Initial Coin Offering), *stock*, *bond*, *investor*, or *sale*. These two narratives equate to 29.2% of all cryptocurrency-related news. The second-largest narrative, producing 12.5% of all cryptocurrency-related news, describes the technical or newly established business that has been formed around the technology. In this case, words such as *blockchain*, *technolog*, *startup*, *venture* (most likely referring to venture capital) make up this topic. The next two narratives describe regulatory themes. The first of these is orientated to political legislation: *rule*, *administr*, *polici* or *congress* are among the most representative words here while the second

4 For more details about the implementation, see [18].

5 Note that the log-likelihood approach [16] retrieved 40 as the optimal number of topics. I decided to go for 10 topics for two reasons: firstly, interpretability of the topics (which depends on the words that compose them) was not higher when using 40 topics than 10; and, secondly, given that I am interested in broader narratives it is more convenient to use fewer topics (as opposed to using many topics that I then group into common themes), as long as interpretability is not an issue.

describes banking regulation; words such *regul*, *launder*, *trade*, *rule* or *tax* frame this narrative. Finally, we find two security-crime orientated narratives: *hacker*, *ransomwar*, *secur*, *breach*, *cybersecur*, *arrest*, *crimin* or *lawyer* being among the words characterising these topics. It is worth noting that there are three additional narratives that were not selected since they do not fall in any of the four categories of interest.[6]

Note that these words do not say anything that indicates the sentiment. A new line of research combines sentiment analysis with topic modelling to account for the tone of the narratives (see [19] or [20]). Following Larsen and Thorsrud [20], I build each narrative-sentiment time series in a few simple steps. Firstly, I find the sentiment in each news-article using *TextBlob*, a publicly available library for natural language processing.[7] *TextBlob* goes beyond simply counting negative vs. positive words in an article by taking into account negation (e.g. *not great* will be rightly assessed as a negative sentiment) and modifier words (e.g. *very* before *bad* will intensify the sentiment of *bad*). This tool retrieves a measure between -1 (negative sentiment) and 1 (positive sentiment). Secondly, to correctly assess the sentiment behind a particular topic I match the overall-article-sentiment score to the most representative topic in the article.[8] The average sentiment score across topics is illustrated in the third column of Table 1. As expected, the average sentiment score in articles describing investment or technology (0.07 and 0.09, respectively) is higher than the average sentiment score in articles reporting regulatory or security issues (0.065 and 0.04, respectively).

To obtain the time series data, I sum each topic proportion (augmented by its sentiment) per month. This retrieves a measure of the intensity of each topic and its sentiment over time. Finally, given that the total number of articles on the online platform is not constant over time, I divide each time series by the total number of articles containing the word *today* for each month (as the proxy for the total number of articles: see [21]). Note that I merge the two narratives describing investment into one by summing the final time series (after having accounted for the sentiment and the overall number of news articles). The same is done for the two narratives relating to regulation.

3. Methods and data description

3.1. Convergent cross mapping

Convergent Cross Mapping (CMM) is a new causal inference tool developed by Sugihara et al. [5]. CCM is meant for *weakly coupled dynamical systems*: a systems of time series that interact (are coupled) and whose relationships might change over time (are dynamic). In dynamical systems, variables are interdependent and cannot be analysed separately. For this reason, linear approaches such as regression, structural equation modelling, or Granger causality tests render no validity. This is because most linear methods are based on correlations and therefore assume that the system is additive [5]. On this basis, the study of nonlinear dynamical systems requires nonlinear methods that acknowledge state-dependency [6].

Analytical methods have been developed over the recent decade that permit the study of dynamical systems of nonlinear time series data [22–25]. These nonlinear statistical methods are rooted in state-space reconstruction (SSR): lagged coordinate embedding of time series data [26]. SSR does not assume any set of equations governing the system; instead, it recovers the dynamics from the time series. In other words, dynamical systems can be described as the evolution of a set of states over time-based on rules governing the development of states in a high dimensional state-space (i.e. a manifold). Motion on the manifold can be projected onto a coordinate axis, forming a time series. More generally, any set of sequential observations of the system-state (i.e. a function that maps the state onto the real number line) is a time series. Therefore, time series (observations) can be plotted in a multidimensional state-space to recover the dynamics, known as attractor reconstruction [27].

Fig. 1 shows the canonical Lorenz system, a coupled system in x, y, and z where the attractor manifold for the original system is represented by M and the two shadow manifolds which are built using lagged-coordinate embedding (τ in Fig. 1) of x and y are M_x and M_y, respectively. Since x and y are dynamically coupled, points near M_x (the red ellipse in Fig. 1) will correspond at some point to values that are close to M_y (i.e., within the green circle). Namely, the points inside the red ellipse and the green circle have corresponding time indices that enable us to estimate states across manifolds, using y to determine the state of x and vice versa [5].

CCM produces a causal network structure that describes which variables are causally connected, including the direction of causality by using time-delayed embedding of the time series data in a higher dimensional space. Put simply, CCM uses time-delay embedding of one time series to generate an attractor reconstruction and then applies the simplex projection algorithm to estimate current values of another time series. This algorithm predicts the current quantity of one variable (M 1) using the time lags of another variable (M 2) and vice versa. If M 1 and M 2 belong to the same dynamical system

[6] Although the narrative *assets* might resemblance investment, it seems more speculative as words such *say*, *like* or *even* are among the selected words.

[7] See https://textblob.readthedocs.io/en/dev/.

[8] To illustrate this last step, imagine that we have a news-article with the following topic composition: 80% *investment* and 20% *security*. Let the overall sentiment of this article be very positive (e.g. it is describing huge gains of investors in the cryptocurrency market). Since we want to match the overall sentiment of the article to the topic *investment*, we first classify that article according to its most representative topic. If we do not do this, both topics (*investment* and *security*) will be allocated a positive sentiment.

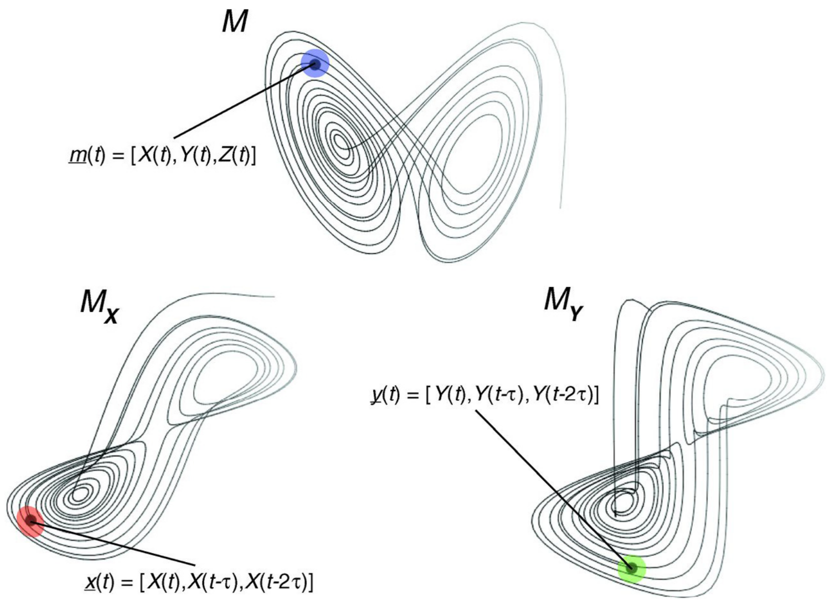


Fig. 1. Lorenz system.
Source: Sugihara et al. [5].

(i.e., they are causally linked), the cross-mapping between them shall be "convergent". Convergence means that the cross-mapping skill (ρ) improves as the library size increases. Recall that the library is the training portion of the data that will be used to reconstruct the manifold. To this end, the more data in the library, the denser the reconstructed manifold, and greater the accuracy of the prediction (i.e., the simplex projection). Therefore, CCM relies on two elements to obtain causality, as follows.

Cross-mapping:
If the time series data from each variable, say x and y, can be used to obtain the shadow manifolds M_x and M_y that are approximations to the true attractor. In other words, these two variables are connected because they are part of the same dynamical system given that they both represent a dimension in the state-space.

Convergence:
If x causes y, then the estimate of x obtained from M_y should improve as the number of points sampled from M_y becomes larger. This is because the library of samples will become a more accurate representation of the attractor, and the nearest neighbour points will be closer and closer to y_t.

3.2. Data pre-processing and description

I proxy cryptocurrency prices according to the exchange rate between the US dollar and Bitcoin (using a natural logarithmic scale) which is obtained from Coindesk.[9] As the leader of cryptocurrencies, Bitcoin prices strongly correlate with other major cryptocurrency prices and, most importantly, have been available for a longer period of time. This allows me to stretch the time period as much as possible.

Following the recommendations of Chang et al. [6], I pre-process the time series data in two steps. Firstly, I remove any linear trends of each time series (prices and narratives) using regression.[10] Secondly, each time series is normalised to zero mean and unit variance; in this way I ensure that all variables have the same level of magnitude for comparison and avoid constructing a distorted state-space. It is worth noting that the frequency of the time series is monthly. I have chosen to use monthly data as there are a lot of missing observations in the narratives; that is, articles concerning cryptocurrencies were not written on daily basis.[11] This is especially the case in data from the first months of Bitcoin existence. Any vector containing missing data is also omitted during computation. Therefore, missing data impart an

9 See www.coindesk.com.

10 Alternatively, one could take the first differences of each time series to guarantee stationarity. However, taking the first differences would remove information relating to any long-run relationship between the series [28]. For this reason, I prefer to use a regression approach in order to guarantee stationarity.

11 Note that only 62% of the days in our sample contain articles written about cryptocurrencies.

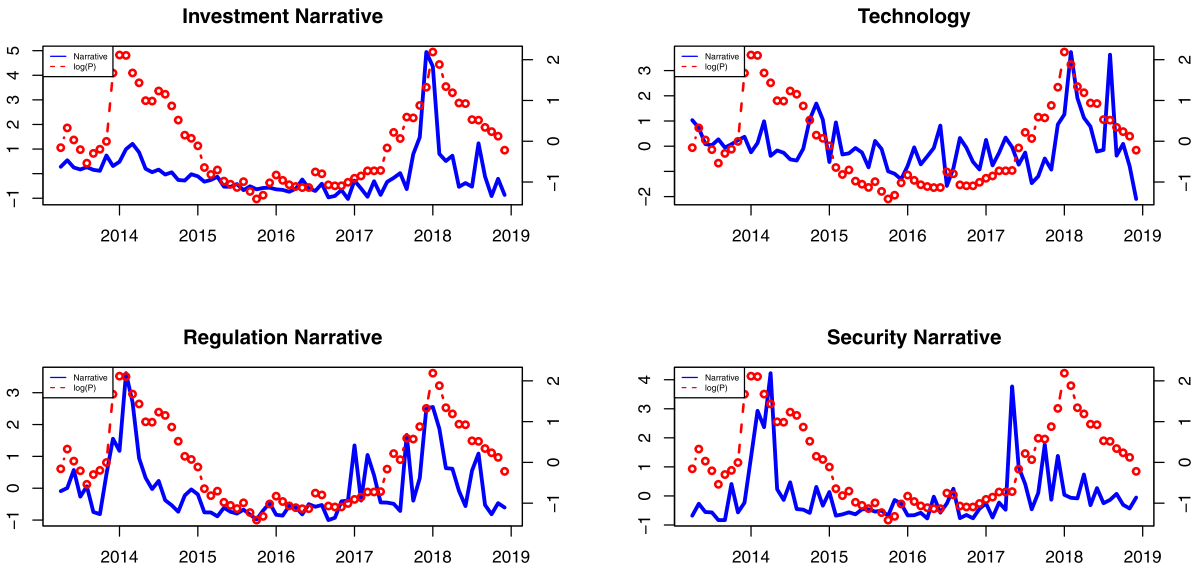


Fig. 2. Narratives and prices. **Notes:** Solid blue lines correspond to the four narratives unveiled by the LDA algorithm (the left-hand legend) while red dotted line correspond to the natural logarithm of Bitcoin prices (the right-hand legend). All of the series are linearly detrended using regressions and the outcome is standardise to mean 0 and unit standard deviation.

unavoidably negative influence on the performance of CCM [6]. In addition, I find no periodicity nor a cyclical component in the time series. This is important, since failing to account for strong seasonality when it exists will produce distortions in the manifolds (see [29]).

Fig. 2 shows the evolution of monthly prices and narratives from April 2013 to December 2018. Overall, we can see a co-movement between the evolution of these narratives and prices. This is especially the case during the two sharpest rises in Bitcoin prices (in early 2014 and end of 2017). However, an important distinction arises: while investment and technological narratives display the highest peak during the second sharpest rise in prices (at the end of 2018, when prices almost reached $20,000) this is not the case for the two most dismissive narratives (Regulation and Security). We observe that the Security narrative barely shows any increases during this time; however, it does when Bitcoin prices stagnate. The biggest spike for the Security narrative occurs during February 2014; this was when Mt. Gox, the world-leading Bitcoin exchange at the time, announced that 85,000 Bitcoin belonging to customers were missing. During this month, Mt Gox suspended trading, closed its exchange service, and filed for bankruptcy protection from its creditors.[12] Given the legal implications, it is not surprising to see a spike in the Regulation narrative also during this month. The second-biggest peak in the regulation narrative occurred in December 2017, when Bitcoin futures were launched thanks to the Gain regulatory approval. The second-biggest spike in the security narrative takes place in May 2017, when the exchange Binance, reported that 7000 Bitcoin were stolen; this revelation caused Bitcoin prices to drop by around 5%.[13]

Before formally testing any pair-wise causal relationship via CCM, I briefly present the optimal embedding dimensions of the variables used for the manifold reconstruction. The embedding dimensions are equivalent to the lags used for the reconstruction of the manifold. Failing to find the optimal number of embeddings will result in poorly reconstructed states. If the number of embeddings falls short, reconstructed states will overlap, causing it to appear to be the same even though they are not [30]. This, in turn, will result in poor forecast performance because the system behaviour cannot be uniquely determined in the reconstruction. Therefore, to find the optimal number of embedding dimensions, it is common to rely on the prediction skill [30]. Following Sugihara and May [22], I use the Simplex Projection which uses the nearest-neighbour forecasting method.[14] As can be seen in Panel A of Fig. 3, the optimal number of embeddings varies across time series, suggesting that the dynamics of the system might be high dimensional [30].

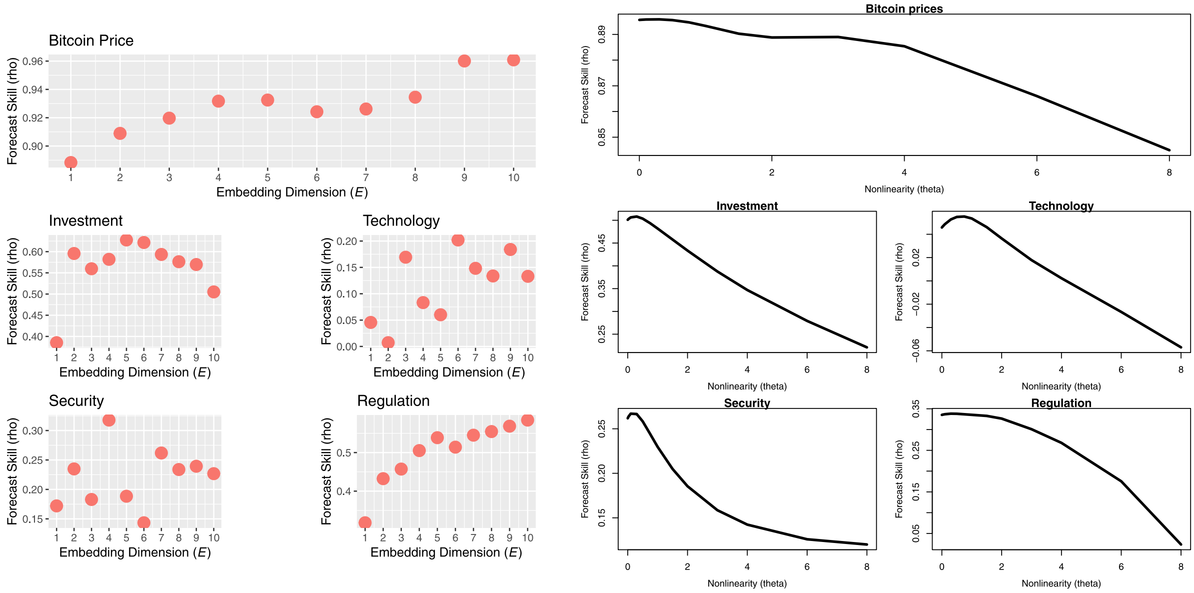
Using these optimal embedding dimensions, we can identify any *nonlinearity* in the system. I do so by using the S-maps function,[15] which applies the nonlinear tuning parameter Θ to determine the strength of the weighting when fitting the local linear map. As can be seen in Panel B of Fig. 3, there is an initial rise in the forecast skill when $\Theta > 0$ and a consequently drop. This is indicative of nonlinear dynamics as allowing the local linear map to vary in state-space produces a better description of state-dependent behaviour [30].

12 See https://www.ft.com/content/6636e0e8-a06e-11e3-a72c-00144feab7de#axzz2v8w0y2mI.

13 See https://www.coindesk.com/hackers-steal-40-7-million-in-bitcoin-from-crypto-exchange-binance.

14 To identify the optimal embedding dimension E for each standardised time series, I use the function *simplex()* Ye et al. [30].

15 See *s_map*() function of the *rEDM* library Ye et al. [30].



(a) Optimal embedding (b) Nonlinearity as a function of Forecast skill

Fig. 3. Optimal embedding dimension and nonlinearities as a function of forecast skill.

4. Empirical results

4.1. Baseline results

Fig. 4 shows the results when applying CCM to Bitcoin prices and different narratives. Overall we observe bi-directional causal relationships between narratives and cryptocurrency prices. That is, price dynamics influence the propagation of news-articles describing the cryptocurrency phenomenon while simultaneously, narratives influence price dynamics. However, the strength of causal relationships depends strongly on the narrative. As such, results suggest that cryptocurrency prices promote news characterising investment and regulation while not promoting those describing technology or security issues. This can be explained by the fact that price changes directly affect investment, at the same time putting pressure on policymakers to adopt new regulations. For example, increases in prices will signal higher adoption levels, as a result of which regulatory institutions might be more prone to acting.

We also observe that the investment narrative affects price dynamics, although the strength of this effect is lower than that from prices to narratives: as the library size increases, so does the cross-mapping skill (the property of convergence); however, values at the end of the library size are further from 1 than in the opposite direction of the causality. This seems to indicate that the press acts as a signal booster of events related to investments, that is, it reacts to price dynamics by describing the investment side. This increase in investment-related news will auspicious further price changes. In addition, we also perceive a causal effect from the regulation narrative to prices. This is not surprising, since Kristoufek [31] has already documented that regulation from China has had a negative impact on prices. Regarding the causal effects between prices and the technological or security narrative, the results are hard to interpret. The technological narrative seems to affect prices more than the other way around,[16] but values for the cross-map skill remain very low. Finally, prices do not seem to affect the security narrative (i.e., there is no property of convergence), while the results the other way around (the effect of the security narrative on prices) might be not statistically significant (with low values in the cross-map skill). For this reason, we need to test whether these results are statistically significant.

Endorsed by Ye et al. [30], I use the randomisation tests with a surrogate time series to assess whether or not these causal effects are significant. This test compares the output produced by the CCM (cross map skill as a function of the library size) for the actual model and an alternative model generated through a surrogate time series under different null models (see Fig. 6 in the Appendix).[17] Confirming suspicions, I observe weak significance in the causal link between the technological narrative and prices at the 90% confidence level. This is because the cross-map skill of the actual model is fairly close to that produced by the surrogate one for different library sizes. The same occurs for the security narrative.

16 In both cases we observe convergence and a stronger causal link from narratives to prices than the other way around.

17 In order to know whether the recovered information about X is unique to the real data rather than just a statistical property of Y we generate surrogates of Y. We then compute cross mapping from surrogates of Y to the actual X from the null distribution of multiple surrogates to pull the 90% quantile for testing significance.

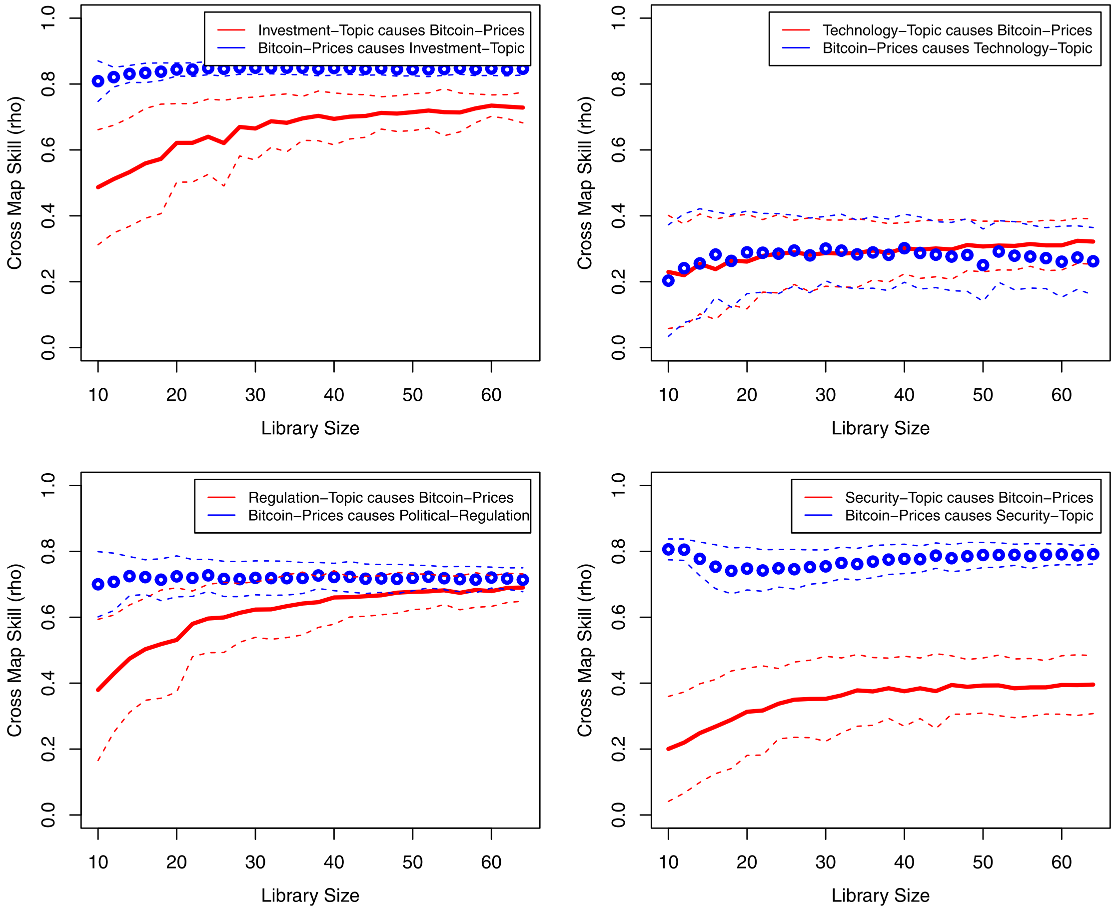
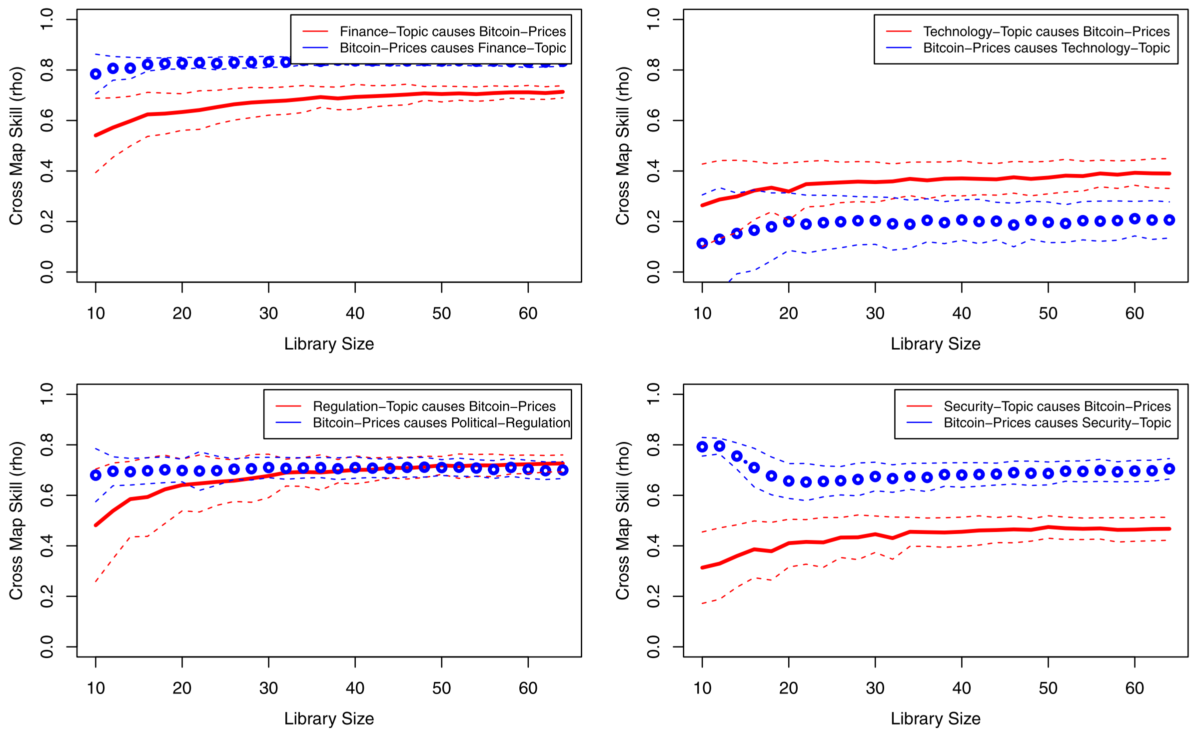


Fig. 4. Convergent Cross Mapping results between narratives and Bitcoin prices. Correlation coefficient (y-axes) as a function of the library size (x-axes). **Notes:** Discontinuous lines represent two standard deviations. In the conventional labelling of CCM 'prices cross-map Topics' is interpreted as 'Topic causes prices'.

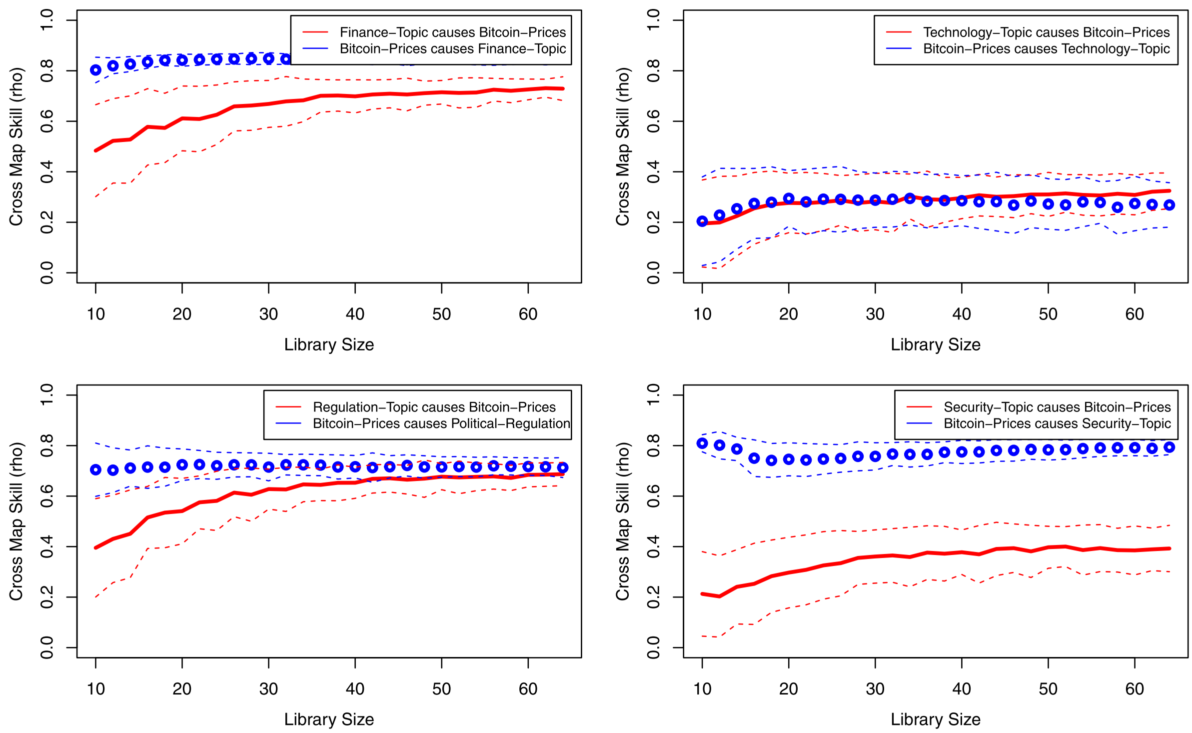
A natural way to contrast these results is to use conventional methodologies such as the Granger causality (GC) test. Although, as earlier explained, this test assumes the separability of the system and is, therefore, not suited to complex dynamical ecosystems, it is worth comparing the results from both methodologies. Given that the GC test is very sensitive to the number of lags chosen, I present results for 3, 6, and 12 lags. As can be seen in Table 2 in the Appendix, results retrieved by the pairwise GC tests slightly resemble those of CCM. The only narrative that Granger-causes prices is the investment narrative. However, this test does not acknowledge that prices Granger-cause the investment narrative. In addition, while prices Granger-cause the regulation narrative, no causality is shown the other way around. Lastly, and perhaps more shockingly is the fact that prices Granger-cause the security narrative.

4.2. Time-delayed causal interactions

CCM results are only valid if the ecosystem does not suffer from "generalised synchrony". Generalised synchrony occurs when there is an exceptionally strong unidirectional force in the ecosystem. In its presence, the dynamics of a response variable, say variable y, become dominated by those of the driving variable, call it x, in such a way that the full system (consisting of both the response variable and the driving variable) collapses to just that of the driving variable [32]. Although there is no causal effect of y on x, the states of the driving variable x can uniquely determine the response variable y; as a result, the CCM causal effect will be observed in both directions. Thus, CCM causal effects will be limited by the fact that CCM may not be able to distinguish between bidirectional causality and strong unidirectional causality that leads to synchrony [32]. For this reason, and to account for any possible delay effect of narratives on prices and vice versa, I consider different lags for cross mapping. Given that we are dealing with monthly data, I opt for for one lag, that is, a month forward or backward from the narratives to prices. As can be seen in Fig. 5, results when prices are lagged one month (Panel a), or the narratives are lagged one month (Panel B) strongly resemble those of the baseline specification. Therefore, there is no reason to believe that there is "generalised synchrony" in this complex ecosystem.



(a) Prices at $t+1$



(b) Narratives at $t+1$

Fig. 5. CCM results for different time delay values.

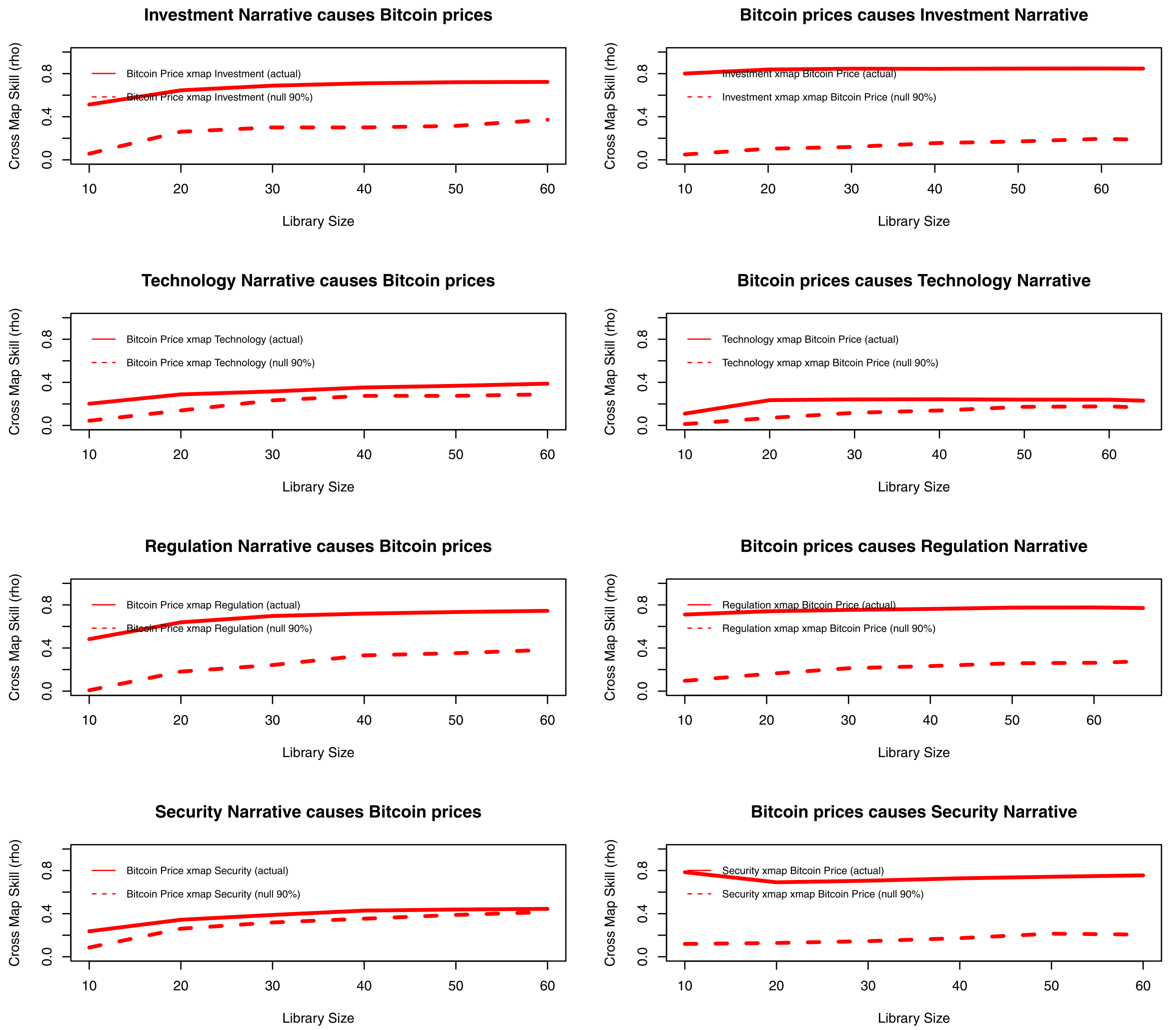


Fig. 6. Test of significance of the baseline results. **Notes:** To test for the significance of cross map effects, I use randomisation tests with surrogate time series.

5. Conclusion

In this paper, I explore the causal relationship between the narratives embedded in conventional media-articles describing cryptocurrencies and prices. I unveil two positive narratives (*investment* and *technology*) while two defeatist narratives (*security* and *regulation*). Given that I observe a highly nonlinear behaviour in their time series, I use an innovative tool suited for weakly coupled components of non-linear dynamical systems to establish causal links: Convergent Cross Mapping (CCM).

Results indicate a strong bi-directional causal relationship between narratives and cryptocurrency prices: price dynamics influence the propagation of news-articles describing the cryptocurrency phenomenon while, simultaneously, narratives influence price dynamics. However, while cryptocurrency prices appear to have a strong impact on news regarding investment and regulation, they do not appear to influence news relating to technology or security issues. This can be explained by the fact that price changes directly affect investment as well as putting pressure on policymakers to adopt new regulations. Price dynamics should not influence criminal activity or technological innovations in the short run. I identify a strong causal link from the investment and regulation narratives to prices, while the technology and security narratives have an effect on prices (albeit it is a weak causal link) but not the other way around. This makes sense given that technological innovations or security threats should not depend on prices.

Overall this work demonstrates that there is a link between narratives and cryptocurrency prices. Stories do motivate and connect people through their values and needs, which ultimately have an impact on economic agents.

Table 2
Granger causality tests.

	Direction of causality	Investment narrative	Technology narrative	Regulation narrative	Security narrative
Prices (3 lags)	←	0.067*	0.708	0.707	0.635
Prices (3 lags)	→	0.226	0.064*	0.001***	0.038**
Prices (6 lags)	←	0.142	0.954	0.381	0.742
Prices (6 lags)	→	0.271	0.199	0.003***	0.004**
Prices (12 lags)	←	0.024**	0.763	0.539	0.13
Prices (12 lags)	→	0.329	0.353	0.050*	0.168

Notes: *P*-value reported. Monthly data used. Prices refer to the natural logarithm of Bitcoin prices. In the lines in which the direction of causality is ←, the null hypothesis is that the corresponding narrative does not Granger-causes Bitcoin prices. When the direction of causality is →, it is the other way around.
*Significant at 10%.
**Significant at 5%.
***Significant at 1%.

Appendix. Additional figures

See Fig. 6 and Table 2.

References

[1] J. Detrixhe, Irrational Exuberance, (05 September), QUARTZ. Availabe at, https://qz.com/1067557/robert-shiller-wrote-the-book-on-bubbles-he-says-the-best-example-right-now-is-bitcoin/, 2017.
[2] G.A. Akerlof, D. Snower, Bread and bullets, J. Econ. Behav. Organ. 126 (2016) 58–71.
[3] R.J. Shiller, Narrative economics, Amer. Econ. Rev. 107 (2017) 967–1004.
[4] A.M. Antonopoulos, The Internet of Money, first ed., Merkle Bloom LLC, ISBN: 1537000454, 2016.
[5] G. Sugihara, R. May, H. Ye, C. Hsieh, E. Deyle, M. Fogarty, S. Munch, Detecting Causality in complex ecosystems, Science 338 (6106) (2012) 496–500.
[6] C.-W. Chang, M. Ushio, C. Hsieh, Empirical dynamic modeling for beginners, Ecol. Res. 32 (6) (2017) 785–796.
[7] J.C. Go, L.H. Ederington, Is a bond rating downgrade bad news, good news, or no news for stockholders?, J. Financ. 48 (5) (1993) 2000–2008.
[8] A.N. Bomfim, Pre-announcement effects, news effects, and volatility, monetary policy and the stock market, J. Bank. Financ. 27 (1) (2003) 133–151.
[9] A. Gande, D.C. Parsley, News spillovers in the sovereign debt market, J. Financ. Econ. 75 (3) (2005) 691–734.
[10] L. Kristoufek, Bitcoin meets google trends and wikipedia, quantifying the relationship between phenomena of the internet era, Sci. Rep. 3 (3415) (2013).
[11] D. Garcia, C.J. Tessone, P. Mavrodiev, N. Perony, The digital traces of bubbles, feedback cycles between socio-economic signals in the bitcoin economy, J. R. Soc. Interface 11 (99) (2014).
[12] A. Yelowitz, M. Wilson, Characteristics of bitcoin users. an analysis of google search data, Appl. Econ. Lett. 22 (13) (2015) 1030–1036.
[13] R.C. Phillips, D. Gorse, Mutual-excitation of cryptocurrency market returns and social media topics, in: 3rd International Conference on Knowledge Engineering and Applications (ICKEA 2018) –Moskow, Russia, 2018.
[14] S. Begušić, H.E. Kostanjčar, B. Podobnik, Scaling properties of extreme price fluctuations in bitcoin markets, Physica A 510 (2018) 400–406.
[15] D.M. Blei, A. Ng, M.I. Jordan, Latent Dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[16] T. Griffiths, M. Steyvers, Finding scientific topics, Proc. Natl. Acad. Sci. USA 101 (1) (2004) 5228–5235.
[17] M.D. Hoffman, D.M. Blei, F. Bach, On-line learning for latent Dirichlet allocation, Neural Inf. Process. Syst. (2010).
[18] R. Řehůřek, P. Sojka, Software Framework for Topic Modelling with Large Corpora, in: Proc. LREC 2010 Workshop on New Challenges for NLP Frameworks, 2010.
[19] B. Saltzman, J. Yung, A machine learning approach to identifying different types of uncertainty, Econom. Lett. 171 (2018) 58–62.
[20] V.G. Larsen, L.A. Thorsrud, Business Cycle Narratives, CESifo Working Paper Series 7468, CESifo Group Munich, 2019.
[21] A. Azqueta-Gavaldón, Developing news-based economic policy uncertainty index with unsupervised machine learning, Econom. Lett. 158 (2017) 47–50.
[22] G. Sugihara, R.M. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in a data series, Nature 344 (1990) 734–741.
[23] C.N.K. Anderson, C.H. Hsieh, S.A. Sandin, R. Hewitt, A. Hollowed, J. Beddington, R.M. May, G. Sugihara, Why fishing magnifies fluctuations in fish abundance, Nature 452 (2008) 835–839.
[24] S.M. Glaser, M.J. Fogarty, H. Liu, I. Altman, C.-H. Hsieh, L. Kaufman, A.D. MacCall, A.A. Rosenberg, H. Ye, G. Sugihara, Complex dynamics may limit prediction in marine fisheries, Fish Fish 15 (2014) 616–633.
[25] H. Ye, R.J. Beamish, S.M. Glaser, S.C.H. Grant, C.H. Hsieh, L.J. Richards, J.T. Schnute, G. Sugihara, Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling, Proc. Natl. Acad. Sci. USA 112 (2015) E1569–E1576.
[26] F. Takens, Detecting strange attractors in turbulence, in: D.A. Rand, L.S. Young (Eds.), Dynamic Systems and Turbulence, Springer, New York, 1981, pp. 366–381.
[27] N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Geometry from a time series, Phys. Rev. Lett. 45 (1980) 712–716.
[28] C. Brooks, Introductory Econometrics for Finance, second ed., Cambridge University Press, New-York, 2014.
[29] E.R. Deyle, M.C. Maher, R.D. Hernandez, S. Basu, G. Sugihara, Global environmental drivers of influenza, Proc. Natl. Acad. Sci. USA 113 (46) (2016) 13081–13086.
[30] H. Ye, A. Clark, E. Deyle, G. Sugihara, rEDM, An R package for Empirical Dynamic Modeling and Convergent Cross Mapping, https://www.cran.r-project.org/web/packages/rEDM/rEDM.pdf, 2018.
[31] L. Kristoufek, What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis, PLoS One 10 (4) (2015) e0123923.
[32] H. Ye, E.R. Deyle, L.J. Gilarranz, G. Sugihara, Distinguishing time-delayed causal interactions using convergent cross mapping, Sci. Rep. 5 (14750) (2015).