

INTRODUCTION

Blockchain technology and innovation in financial systems have led to the birth of an alternative mode of currency, called cryptocurrencies. Cryptocurrencies have emerged as a digital currency which offers a decentralized and transparent financial system, as opposed to the existing centralized traditional financial system. They provide an easy, fast and seamless peer to peer transfer (Corbet et al., 2019). One of the most popular cryptocurrency is Bitcoins (Briere, Oosterlinck and Szafarz, 2015), which has emerged as a form of alternative investment for portfolio diversification as well. Frequent high fluctuations in Bitcoin prices have made it an extremely volatile investment. Price movements in an interval of $\pm 8000\%$ have been observed in Bitcoins (Ciaian, Rajcaniova and Kancs, 2016) and studies have also have emphasized the need to examine its price volatility (Dyhrberg, 2016; Katsiampa, 2017). Hence, it becomes imperative to explore the nature of Bitcoin price fluctuations to reduce the risk of investment in Bitcoins.

Mechanisms to understand Bitcoin market price fluctuations are very complicated. Despite receiving tremendous research attention, it is still surrounded by a lot of scepticism. It is not backed by any state, nor does its fundamental value reside on any underlying asset (Bariviera et al., 2017). Moreover, the price of Bitcoin today has increased to as much by 5000% since its initial price in 2011 (Urquhart, 2016).

Exploring the statistical properties of Bitcoin prices within the gamut of market efficiency (Fama, 1970), also known as the efficient market hypothesis (EMH), has become a debatable issue among scholars.

Market efficiency implies that all available information is fully reflected in the market prices and hence, markets are subjected to a random walk model. While EMH makes room for short term memory persistence in the market, it rejects the presence of long term memory persistence in the financial time series of tradable assets. If there is long term memory, it will imply that traders can develop a profitable trading strategy as markets do not follow a random walk model (Briere, Oosterlinck and Szafarz, 2015).

Market irregularities enable to develop predictable price formations which are exploited by the investors till it is discovered by the market (Henrique, Sobreiro, & Kimura, 2018). Scholars have made several attempts to disprove the EMH by examining the irregular movements in market price series (Joseph and Mazouz, 2010; Tripathy, 2018), leading to the existence of profitable trading strategies. While (Fama, 1970) assumed a linear stochastic process in assuming markets to follow a random walk model, scholars have urged to explore the nonlinear stochastic process for the same (Lim and Brooks, 2011). Exploring the linear and nonlinear stochastic movements in Bitcoins, scholars have attempted to understand the irregularity in Bitcoin prices which can lead to better prediction of its movement (Ramirez, Rodriguez, & Valdez, 2018).

With no government back up and fundamental tangible value, Bitcoin has become very speculative and sensitive to any fake or even real news dealing with economic or political events (Ramirez, Rodriguez, & Valdez, 2018). However, economists have been comparing Bitcoin similar to gold as both assets are limited in availability along with being difficult to mine (Dyhrberg, 2016). It has been studied that gold prices follow a cyclical process and different factors influence gold prices differently. Examining movements in gold prices as a multi-scale problem, (Wen et al., 2017) looked at the irregularity at gold prices to find that gold price series are multi-scale in nature. A multi-scale situation arises when time series consists of components which are at disparate scales and requires analysis from different levels of resolution (e.g., hourly, weekly etc.). The dynamics of the series at these resolutions also varies, thereby affecting the inferences of the underlying process. Hence multi-scale series warrants decomposition (Nawrotha and Peinke, 2006) and aggregation (Drost and Nijman, 1993) of the original series at different time resolution.

While few countable studies on Bitcoins have shown Bitcoins to be subjected to market inefficiency and moving towards efficiency (Kristoufek, 2018; Zargar and Kumar, 2019), none of these studies has examined the irregular movements in Bitcoin price series as a multi-scale problem. The study intends to

investigate the multi-scale nature of the Bitcoin price series. Exploring multi-scale movement features of Bitcoin price series can also lead to better prediction of Bitcoin prices.

EMH states that on account of markets following a random walk model, the future market movements cannot be predicted consistently. However, computationally intensive techniques employing machine learning algorithms are now being extensively used to develop profitable stock trading strategies. Multiple studies are being done to examine whether machine learning tools can be leveraged to anticipate stock predictions in a consistent manner (Patel et al., 2015; Nardo et al., 2016; Nelson et al., 2017). Studies have also explored machine learning algorithms to predict Bitcoin prices (Ciaian, Rajcaniova and Kanacs, 2016; Jang and Lee, 2018; McNally, Roche and Caton, 2018). However, no research has considered the multi-scale property of Bitcoin prices to assess whether the price series can be predicted. Being a highly volatile currency and subjected to a lot of ambiguity, it becomes more interesting to find whether machine learning algorithms can be used to effectively predictive the power of cryptocurrency. The objective of this study is thus twofold. Firstly, it examines the nature of Bitcoin price series and assesses their multi-scale properties. Decomposing Bitcoin price series using an algorithm is effective if it enables to identify each physical mode component belonging to different cycles. While variants of Auto-Regressive Conditional Heteroscedasticity (ARCH) models and wavelet decomposition are widely used to identify multi-scale nature properties of volatile series, scholars have subjected them to many shortcomings. While ARCH models are highly model specific (Shik Lee, 2004), the decomposition scales are not adjustable in wavelet analysis (Wen, Yang, Gong, and Lai, 2017). Empirical mode decomposition (EMD) techniques generate a more adaptable decomposition for non-stationary series (Huang et al., 1998). Hence, this study proposes to assess the multiscale properties of Bitcoin price series by using EMD and extended versions of EMD algorithms (Torres, Colominas, Schlotthauer, and Flandrin, 2011) like empirical ensemble mode decomposition (EEMD) and complete empirical ensemble mode decomposition (CEEMDAN). Secondly, it uses a computationally intensive support vector machine learning (SVM) algorithm to examine whether Bitcoin prices can be predicted. Scholars have proposed SVM based regression models as a promising alternative to financial time series prediction methods by giving more optimal results (Kim, 2003; Lahmiri, 2013; Peng et al., 2018).

BACKGROUND

Cryptocurrency markets have received phenomenal investor attention and growth since the last few years. Being relatively young, only a handful of studies have been done so far on the cryptocurrency market. With Bitcoins, being the most popular cryptocurrency, myriad studies have explored the nature of Bitcoin price series from multiple perspectives. One set of studies have investigated the market efficiency of Bitcoin prices. Another set of studies have tried to examine the determinants of Bitcoin prices along with examining its capabilities as a financial asset. Corbet et al., (2018) found cryptocurrencies to be isolated from other financial assets and provide diversification in an investor portfolio.

Using daily Bitcoin price series from March 2013 till July 2017 (Bartos, 2015), examined the EMH to find the effect of public announcements on Bitcoin prices. They found that Bitcoins follow EMH and reflect all available information in the market. Hence, it is not possible to develop prediction model for profitable trading strategies unless the investor has access to insider information. The study also found that global macroeconomic indicators and investor speculations through Wikipedia searches do not have a significant influence on Bitcoin prices. However, refuting the results of Bartos (2015), another study using machine learning algorithm of support vector machine (SVM), found that Bitcoin prices have a significant relation with macro variables and collective mood sentiments sourced through Twitter (Georgoula, Pournarakis, Bilanakos, Sotiropoulos, and Giaglis, 2015). Using monthly Bitcoin returns from 2011-2016, Zhu, Dickinson and Li (2017) built a VECM model to examine the influence of macroeconomic factors and gold price on Bitcoin returns. They found the US dollar to have the most significant impact on Bitcoin returns and gold prices having the least. Applying the GARCH framework

applicable to finance research, Dyhrberg (2016) examined the Bitcoin returns in comparison to a commodity, i.e. gold prices and a currency, i.e. dollar-euro exchange rates. The results showed Bitcoins having similarities to dollar and gold, and hence it can be treated both like a currency and a commodity.

Studying daily Bitcoin returns from 2010 till 2016, Urquhart (2016) analyzed the market efficiency of Bitcoins by using multiple robust time series tests including run tests, Bartels test, variance ratio test, automatic variance test, BDS test and Hurst exponent test. The results showed Bitcoin returns not being weakly efficient over the entire sample period but suggested the returns to be moving towards market efficiency. Using a power transformation of Bitcoin returns, Nadarajah and Chu (2017) used different tests as done by (Urquhart, 2016) to show that Bitcoin returns indeed follow the EMH. Apart from the tests used by (Urquhart, 2016), their study also used spectral shape tests, generalized spectral test and portmanteau test. Leveraging fractional integration and R/S analysis, Caporale, Gil-Alaña, and Plastun (2018) found evidence of long term memory persistence in cryptocurrencies. Refuting the EMH, their study showed the cryptocurrency market to be inefficient, leading to the generation of abnormal profits.

Exploring stylized facts of Bitcoins market, Bariviera et al., (2017) found Bitcoin returns to change dynamically over time with the presence of long term memory dependence. Founding more evidence against EMH in cryptocurrency markets, Caporale and Plastun (2017) found Bitcoin returns subjected to day of the week effect. Examining the time-varying nature of long term memory dependence in returns, Yonghong, He, and Weihua (2017) found that Bitcoin returns are highly inefficient and do not become efficient over time. Examining the statistical properties of Bitcoin return, Katsiampa (2017) found AR-CGARCH model to explain Bitcoin data best. Moving beyond EMH in Bitcoin markets, scholars also started exploring the adaptive market hypothesis (AMH). The AMH states that market efficiency is time varying in nature as markets evolve over time. Khuntia and Pattanayak (2018) found evolving market efficiency in Bitcoins and validated the presence of AMH in Bitcoin markets. Examining nine different forms of cryptocurrency, Zhang et al., (2018) found all of them to be inefficient markets. Table 1 below gives a summarized overview of a few of the market efficiency studies done on cryptocurrencies, particularly on Bitcoins.

Examining whether Bitcoins provide a safe haven effect and exploring their capability as a financial asset, scholars have examined the hedging properties of Bitcoins. Bouri et al., (2017) examined different hedging capabilities of Bitcoins over different time horizons and found them more suitable as a portfolio diversifier instead of using for hedging. Bouri, Azzi, and Dyhrberg (2017) found lack of asymmetric return volatility relation for Bitcoins along with Bitcoin volatility having an inverse relationship with the US volatility index.

A common and dominant observation has been the validation of market inefficiency of Bitcoins. It implies that profitable trading strategies can be developed as the presence of long term memory has been found in Bitcoin markets. With Bitcoin prices being highly speculative and dynamically changing, it becomes imperative to explore their nature as well. However, no study has been done which explores the multi-scale properties of Bitcoin price data. Exploring the characteristics of Bitcoin price data can enable to better understand their irregular movements. It will also allow examining whether Bitcoin prices can be predicted or not. This study is a step in this direction. The next section explains the methodology used in the study.

Table 1: Representation of the studies done on market efficiency of Bitcoins

Author, Year	Cryptocurrency	Test (Result)	Timeline	Frequency	Methodology
Bartos, 2015	Bitcoin	EMH (Accepted)	2013-2017	Daily returns	Error correction models
Urquhart, 2016	Bitcoin	EMH (Rejected)	2010-2016	Daily returns	Variance test, BDS and Hurst exponent
Nadarajah and Chu, 2017	Bitcoin	EMH (Rejected)	2010-2016	Daily returns	Spectral tests and portmanteau test
Caporale, Gil-Alaña, and Plastun, 2018	Bitcoin, LiteCoin, Ripple, Dash	EMH (Rejected)	2004-2017	Daily returns	R/S analysis, fractional integration
Bariviera, Basgall, Hasperué, and Naiouf, 2017	Bitcoin	EMH (Rejected)	2011-2017	Daily returns	Hurst exponent with detrended fluctuation Analysis
Caporale and Plastun, 2017	Bitcoin, LiteCoin, Ripple, Dash	EMH (Rejected)	2013-2017	Daily returns	ANOVA, the Kruskal-Wallis test, and regression analysis
Yonghong, He, and Weihua, 2017	Bitcoin	EMH (Rejected)	2010-2017	Daily returns	Rolling window approach
Khuntia and Pattanayak, 2018	Bitcoin	AMH (Accepted)	2010-2017	Daily returns	Spectral test in Rolling window approach
Zhang, Wang, Li, and Shen, 2018	Bitcoin, Ripple, Ethereum, NEM, Stellar, Litecoin, Dash, Monero and Verge	EMH (Rejected)	2013-2018	Daily returns	MF-DFA and MF-DCCA

METHODOLOGY

Very volatile price series like gold price and crude oil prices are too complicated to process because of its non-linear and non-stationary characteristics. Hence, studies have attempted to sift the intrinsic modes of the series to analyze such series. Being a financial time series, the Bitcoin price series are first subjected to stationarity tests like unit root tests and BDS (Brock, 1996) to detect for the presence of non-stationarity and non-linearity. Further, as a highly volatile series with multi-scale nature, extracting the complete compositional characteristics is crucial for understanding the dynamics of its price fluctuations. The intrinsic modes functions (IMFs) are the basic components of a time series that also represent the natural characteristics of the series (Arasteh, Moradi, and Janghorbani, 2016). By segregating the different IMF of a series, possible hidden structures of a time series can be extracted (Bouoiyour, Selmi, Tiwari, and Olayeni, 2016).

EMD functions have received critical attention for forecasting from economics and financial literature because of its ability to analyse non-stationary and non-linear series. The important characteristics of EMD are its ability to extract IMFs (sub-series). Using sifting algorithms, IMFs at different granularity of the scale (from fine to coarse) are extracted until the residue becomes a constant. EMD of a time series $x(t)$ is given as in equation (1).

$$x(t) = \sum_{i=1}^{N-1} IMF_i(t) + r_N(t) \text{ ----- (1)}$$

where $IMF_i(t)$ is the element that represents the extracted IMFs at different granularity of the scale, $r_N(t)$ is the residue constant and N the number of nearly orthogonal IMF series extracted.

EMD is best suited when the different IMFs are populated with a uniform scale (Wu and Huang, 2009). EEMD overcomes the issue of scale uniformity across IMFs by averaging out the modes across the ensemble of IMFs. This is done by adding a different series of white noise during each shifting process and averaging out the IMF series over different shifting trials. That is for a given IMF ‘ i ’, the average of the i^{th} IMF over M trials is obtained using the equation (2).

$$IMF_i^{avg}(t) = \frac{1}{M} \sum_{m=1}^M IMF_i^{(m)}(t) \text{ ----- (2)}$$

Although EEMD addresses the issue of scale uniformity, its accuracy is affected due to the addition of white noise. Further, as the number of trails increased its computational complexity also increased. Therefore, CEEMD with adaptive noise smoothing (CEEMDAN) was proposed to provide a much robust interpretation of decomposed components (Torres et al., 2011). Similar to EEMD, but in CEEMDAN a signal to noise ratio parameter is added to control the additional noise supplemented to the residue after the extraction of the first IMF. This unique residue is carried forward over subsequent IMFs and residues extraction. Hence a true IMF can be obtained, and the remaining residue after the extraction of all the IMFs is much reduced. The final residue is given as in equation (3).

$$r_M(t) = x(t) - \frac{1}{M} \sum_{m=1}^M IMF_i^{(m)}(t) \text{ ----- (3)}$$

To identify which decomposition method is most suitable for the Bitcoin price series, the IMFs from EMD, EEMD and CEEMDAN decomposition methods are obtained. The difference between the original Bitcoin price series and the sum of the IMFs obtained through EMD was estimated. This was done for both EEMD and CEEMDAN as well. The decomposition method which resulted in the minimum difference between the Bitcoin price series and the sum of the IMF sequences obtained through decomposition method was selected to give optimal results.

The IMFs from the decomposed price series were aggregated to identify short term, medium term and long term trend in the series. Support vector machines (SVM) based regression models were implemented on the aggregated trend obtained to predict the same. SVM based regression models have shown favourable results for financial time series predictions (Kim, 2003; Wen et al., 2017). Given a set of training data (x_i, y_i) , $i=1, 2, \dots, n$ and $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$, where x_i is the input, y_i is the output and n is the size of the training data, SVM maps the input to a high-dimensional feature space, and a linear model is defined in the high-dimensional space (Tay and Cao, 2001). This linear model is constructed using ‘maximum margin hyperplane’ that is capable of representing non-linear boundaries that exist in the original data. Thus for the given training data, SVM constructs a regression function is given as in equation (4).

$$y = w \varphi(x) + b \text{ ----- (4)}$$

Where $\varphi(x)$ represents the high-dimensional feature space of x . The SVM model developed with the training data is then used on the test data for actual predictions. SVM estimates the coefficients w and b using the principles of linearly constrained quadratic programming. The decision of whether to be hard or soft on the boundaries of separation between data points is decided by a cost parameter. SVM can also construct a non-linear model in the high-dimensional space using different kernel functions namely polynomial and radial functions. Choosing a linear or a non-linear kernel is critical as it decides the complexity and the performance of the SVM model. For time series predictions, the input variables of SVM are a set of consecutive time series that are lagged by 5 days predicting the 6th day price index as given in equation (5).

$$x_n(t) = f(x_{n-1}(t) + x_{n-2}(t) + x_{n-3}(t) + x_{n-4}(t) + x_{n-5}(t)) \text{ ---- (5)}$$

The same approach is adopted, using 15 days and 30 days lagged values for price predictions. The next section presents the data analysis done using the above-mentioned methodology.

DATA ANALYSIS

For the study, daily Bitcoin price data from 3/2/2012 to 3/10/2018 is sourced from <https://in.investing.com>. Figure 1 shows the Bitcoin price series and log of Bitcoin prices graph. From the graph, it can be seen that, though the Bitcoin prices have shown an upward trend, the series has been highly volatile and nonlinear in nature.

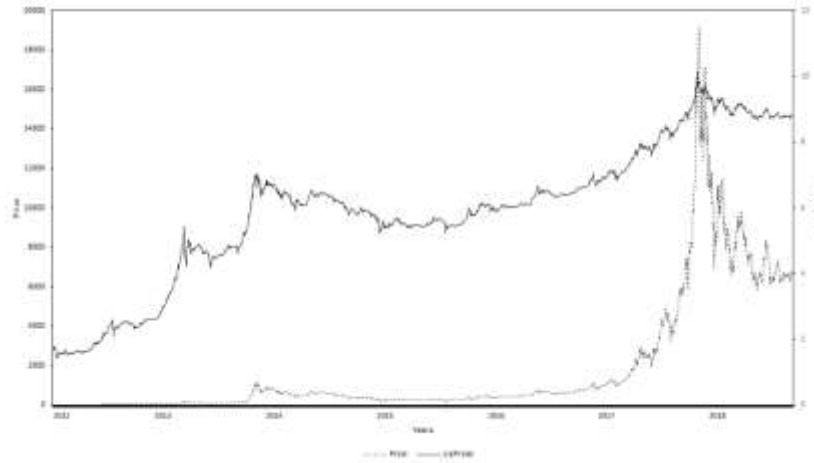


Figure 1 Daily Bitcoin price series in USD and log price series

From the Bitcoin returns graph, it can be observed that the fluctuations in prices are rapid and varying in level. Hence, such price series are suitable for multi-scale level analysis, so as to understand the irregular fluctuations. Also, the graph shows a combination of bullish and bearish phase, with a bullish phase having a long memory. The summary statistics of Bitcoin price series along with stationarity tests are given below in Table 2.

Table 2: Summary statistics of Bitcoin price series

Statistics	Price	Ln(Price)
Mean	1813.404	5.922514
Median	434.90	6.075116
Maximum	19187.00	9.861988
Minimum	4.200000	1.435085
Std. Dev	3194.946	2.092029
Skewness	2.349667	(0.397927)
Kurtosis	8.424186	2.671248
Unit Root tests at level	T statistic / p value	T statistic / p value
Intercept	(1.757567) / 0.4020	(1.475575) / 0.5461
With trend and intercept	(2.773868) / 0.2072	(1.455047) / 0.8445
None	(1.272986) / 0.1875	2.405449 / 0.9965

Notes: For ADF the test critical values at 1%, 5% and 10% level are -3.432864, -2.862537, -2.567346; -3.961846, -3.411670, -3.127711; -2.564920, -1.940955 and -1.616611; for intercept, with trend and intercept, and none respectively.

Unit root tests and BDS tests are performed to examine the non-stationarity and non-linearity in the Bitcoin price series. Unit root tests identified that the Bitcoin price series at level are non-stationary in nature. The unit root tests used to assume the series to have linear dependence among them. BDS test (Brock, 1996) is used to examine the nature of dependence in the Bitcoin price series. The null hypothesis of the BDS test implies the series to be non-stationary with presence of nonlinearity. A predefined distance value ϵ is taken as 0.9 with 6 dimensions in the BDS test by using the fraction of pairs method. The results of the test are reported in Table 3. The null hypothesis is rejected indicating the presence of nonlinearity in the series. Hence, the non-stationary and nonlinear nature of the Bitcoin price series make it suitable to apply decomposition methods.

Table 3: BDS statistics (Fraction of pairs)

M dimension	BDS statistic	P value
2	0.205098	0.0000
3	0.349413	0.0000
4	0.450876	0.0000
5	0.522080	0.0000
6	0.572235	0.0000

The default value of 0 intrinsic mode function (IMF) was used while doing the decomposition analysis. The default value generates the maximum number of IMFs, which in this case came out to be 10 IMFs and a residual value. IMF specific descriptive of the Bitcoin price series is given below in Table 4. The mean period of each IMF is the ratio of the total number of observations to the number of peaks in each period (Ming, Yang, & Cheng, 2016) and it represents the IMF's average time length between the time periods. Peaks are the significant spikes in the given time series and is found as the largest local maxima occurring in the series. The variance percentage explains the IMF's contribution in explaining the total volatility of the series.

From table 4, we can observe that the high-frequency IMF series fluctuate with a lower average period. The variance contribution in explaining the total price volatility is also meagre. On the other hand, the lowest frequency IMFs have longer mean periods with larger variance contribution to the observed price volatility. The dominant IMF is the residual that explains the largest variance of the total volatility. We see that the dominant residual of the EMD method contributes comparatively lesser variance (47%) as with the residuals of the other two methods (77.08% for EEMD and 70% for CEEMDAN). Furthermore, the high-frequency IMFs (IMF1 to IMF 5) of the EMD explain higher variance as compared to the other methods. Overall IMFs of EMD have high fluctuating series explaining more variance than the stable residuals making it least preferred as compared to EEMD and CEEMDAN. Hence, we proceed ahead with only EEMD and CEEMDAN as the decomposition methods to predict the Bitcoin prices.

Table 4: Descriptive of each decomposition method

Series	CEEMDAN			EEMD			EMD		
	Mean Period (days)	Variance	Variance as % of Observed Bitcoin price (Cumulative)	Mean Period (days)	Variance	Variance as % of Observed Bitcoin price (Cumulative)	Mean Period (days)	Variance	Variance as % of Observed Bitcoin price (Cumulative)
IMF 1	2.81	9512.69	0.09% (0.09%)	2.82	9510.78	0.09% (0.09%)	3.16	11103.90	0.11% (0.11%)
IMF 2	3.55	538.81	0.01% (0.10%)	5.20	11243.83	0.11% (0.20%)	6.42	20023.30	0.20% (0.31%)
IMF 3	3.44	8673.32	0.09% (0.18%)	10.10	15125.43	0.15% (0.35%)	14.32	43398.12	0.43% (0.73%)
IMF 4	5.94	10252.12	0.10% (0.28%)	19.80	31209.31	0.31% (0.66%)	32.47	69696.43	0.68% (1.41%)
IMF 5	11.07	31789.03	0.31% (0.60%)	41.98	115053.12	1.13% (1.79%)	65.81	233879.50	2.29% (3.71%)
IMF 6	21.94	53698.81	0.53% (1.12%)	83.97	194865.18	1.91% (3.70%)	173.93	656702.07	6.44% (10.15%)
IMF 7	39.27	113022.00	1.11% (2.23%)	202.92	927813.43	9.10% (12.80%)	405.83	2685538.15	26.35% (36.50%)
IMF 8	90.19	304332.54	2.99% (5.22%)	608.75	774166.70	7.60% (20.40%)	608.75	1335332.60	13.10% (49.60%)
IMF 9	187.31	1351031.99	13.25% (18.47%)	1217.50	497426.43	4.88% (25.28%)		1038955.18	10.19% (59.79%)
IMF 10	811.33	967477.91	9.49% (27.96%)		208405.98	2.04% (27.32%)		3276.53	0.03% (59.83%)
Residual		7131865.49	69.97% (97.93%)		7856705.52	77.08% (104.40%)		4790596.13	47.00% (106.83%)
Variance of the observed Bitcoin price series				10192704					

Two new series were estimated by taking the difference between the original Bitcoin price series and the sum of the sequences obtained by using all the two decomposition methods (EEMD and CEEMDAN). The graph of the two series obtained are compared with the original Bitcoin price series and its difference is shown in Figure 2 and Figure 3 below.

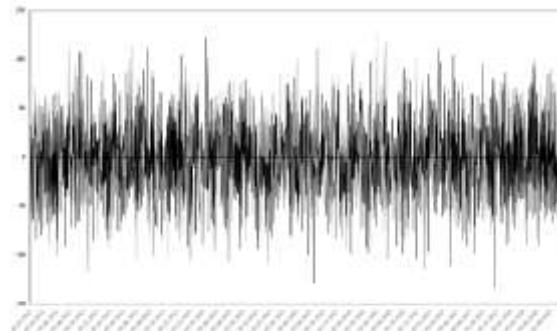


Figure 2 Difference between Bitcoin price series and EEMD

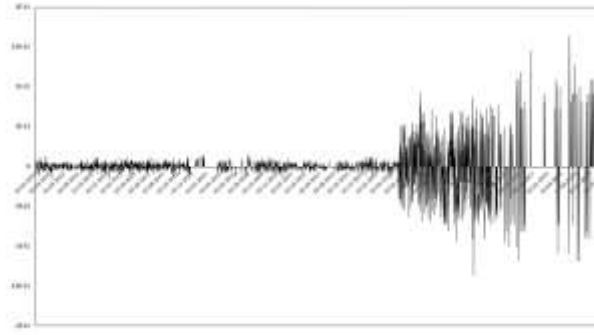


Figure 3 Difference between Bitcoin price series and CEEMDAN

From the graphs, it is evident that CEEMDAN method resulted in the least difference (with difference values concentrated around zero) as compared to the EEMD method. Hence, this paper uses the CEEMDAN method to understand the multi-scale nature of the Bitcoin price series. The analysis was done in R software. Figure 4 shows the graphs of the Bitcoin prices and the trend of the price series obtained through CEEMDAN.

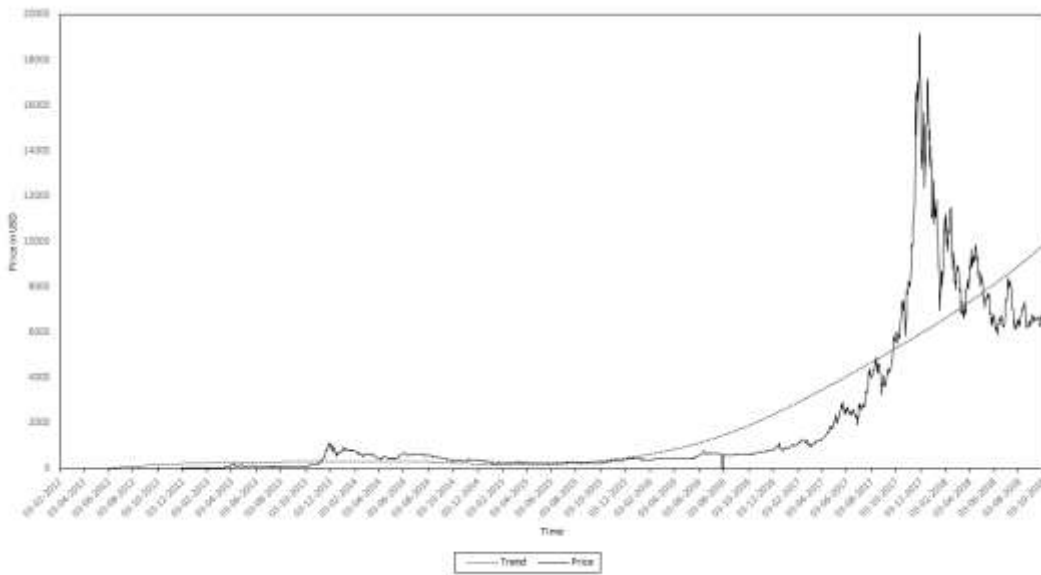


Figure 4 Bitcoin prices and CEEMDAN

The IMFs obtained from CEEMDAN decomposition are further analyzed in detail to aggregate them as short term, medium term and long term trend. Table 5 shows the summary statistics and standard deviation of the IMFs generated from CEEMDAN.

Table 5: Summary statistics of IMF components and residuals

IMF	Mean	Standard Deviation	t statistics for mean
IMF 1	(0.3586)	97.533	(0.18141)
IMF 2	(0.02027)	23.214	(0.04309)
IMF 3	0.3588	93.126	0.19012
IMF 4	(0.6859)	101.235	(0.33423)
IMF 5	(1.4599)	178.781	(0.40354)
IMF 6	(0.1383)	232.745	(0.02933)
IMF 7	(12.776)	334.020	(1.8875)**
IMF 8	(98.01)	548.367	(8.82)***
IMF 9	(144.92)	1,163.560	(6.1459)***
IMF10	172.93	981.204	8.6966***
Residual	1895.5	2,672.030	35.005***
Total		3,192.602	

Note: ** and *** denote significance levels at 5% and 1% respectively

To examine the short term, medium term and long term part of the series, the statistical properties of the IMFs are examined. T test is done for each IMF to assess if the mean of each IMF is significantly different from zero. For IMFs 1-6 it was observed that the mean value is not significantly different from 0. Hence, IMFs 1-6 were grouped together to represent the short term fluctuations in the Bitcoin price series. The t statistic for IMFs 7-10 was statistically significant implying that the mean value is not equal to 0. They were grouped together to represent the medium term fluctuations in the Bitcoin price series. The remaining component, the residual, with a statistically significant t statistic represents the long term part of the Bitcoin price series. Thus the results of the t statistic is used to group IMFs to represent short term, medium term and a long term trend (Wen et al., 2017). Figure 5 shows the graph of the Bitcoin price series and the grouped IMFs together.

From Figure 5 it can be seen that the trend shows the upward long term movement sequence of the Bitcoin price series. From the correlation values also it can be observed that the long term trend, i.e. residual has a highly significant correlation of 0.85 with the Bitcoin price series and accounts for 71.5% of the variance. Post-2016, a steady upward trend can be observed in the Bitcoin prices. The midterm represents the medium frequency which is the medium-price movements and less volatile in nature. The short term represents the high-frequency part of the Bitcoin price series which shows frequent fluctuations in the prices and account for the highest volatility. The cumulative variance of the short term and medium term components are 1.15% and 27.33% respectively.

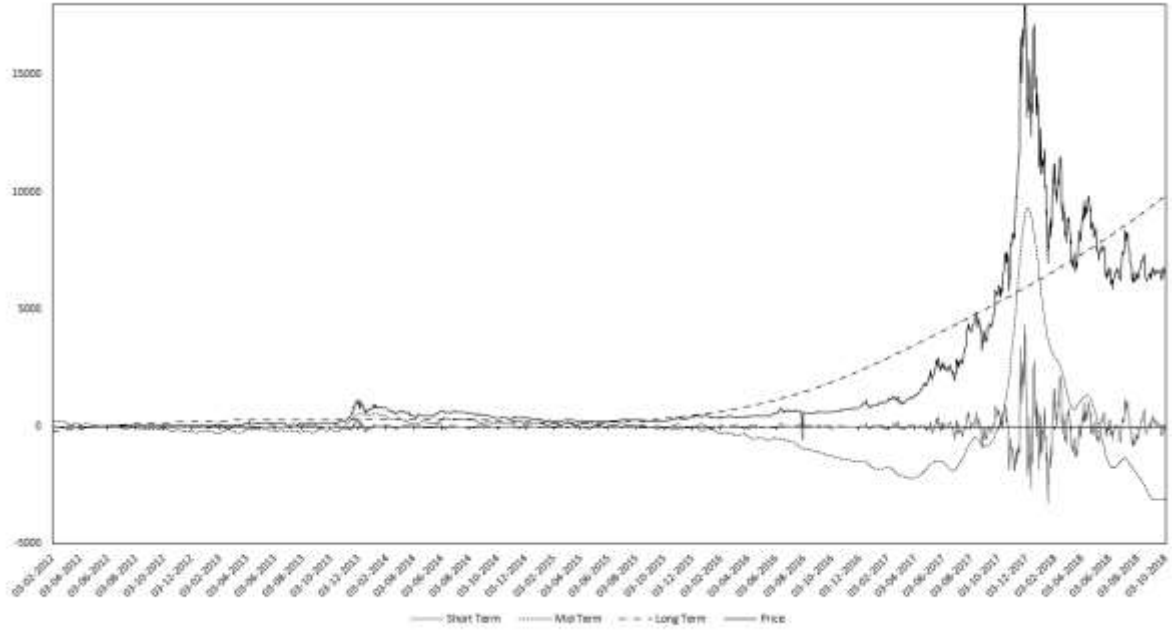


Figure 5 Bitcoin Price IMFs aggregated into short, medium and long term price index

With the presence of multi-scale properties identified through the CEEMDAN decomposition, the study uses the SVM algorithm to assess whether it can help in predicting Bitcoin prices. In this study, we use 2435 daily observation of Bitcoin prices from February 3, 2012 to October 3, 2018. We use the initial 70% of the data (1641 data points) as the training data to develop the SVM regression model, and the latter 30% (794 data points) is used as test data to validate the model. We choose the linear kernel as suggested by Greaves and Au (2015). The best cost parameter is chosen after experimenting with different cost values.

We test the series for 5 days lag, 15 days lag and 30 days lag period. The input variables for price prediction are the indices of the lagged period (Wen et al., 2017). For example, for a 5-day lag period model, the input variables are the price index values of the previous 5 days that are used to predict the 6th-day index. This is followed for all the three series- short, medium and long, of CEEMDAN to predict their respective indices. A total of 9 different (3x3) prediction models are developed for forecasting the price index with 5 days, 15 days and 30 days lag respectively.

The corresponding SVM function of equation (5) can be given as in equation (6).

$$f(x_n(t)) = w_i \varphi(x_{n-i}(t)) + b \text{ ---- (6)}$$

Where $i=1,2,...,5$ for 5 days lag period; $1,2,...,15$ for 15 days lag period and $1,2,...,30$ for 30 days lag period.

We use the R package 'e1071' for SVM modelling (Chang & Lin, 2011). The SVM regression models are trained and tested separately for short, medium and long term series obtained by CEEMDAN. The predicted indices of the three series are then summed up for the respective decompositions to represent the actual Bitcoin price prediction. Finally, the prediction performance of CEEMDAN is evaluated by comparing the actual price values with their predicted values.

Their prediction performance is evaluated using the error metrics Mean Square Value (MSE)- to compare the days' lag of different series and Mean Absolute Percentage Error (MAPE)- to compare the series for a given days' lag. MSE and MAPE can be given as in equations (7) and (8).

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_n(t) - \widehat{x}_n(t))^2 \text{ ---- (7)}$$

$$MAPE == \frac{1}{n} \sum_{i=1}^n \left| \frac{x_n(t) - \widehat{x}_n(t)}{x_n(t)} \right| \text{ ---- (8)}$$

Table 6 below gives the MSE and MAPE of CEEMDAN series.

Table 6: MSE and MAPE for CEEMDAN series

Periods	MSE			MAPE		
	5 days lag	15 days lag	30 days lag	5 days lag	15 days lag	30 days lag
Short Term	146695.50	149909.80	149541.20	1.30	1.35	1.37
Mid Term	0.47	1.32	5.90	0.0005	0.0004	0.0012
Long Term	0.24	0.20	0.28	0.0001	0.0001	0.0001
Combined predicted series	146699.96	149703.20	149542.48	0.0316	0.0321	0.0323

Based on MSE values, from table 6 we can observe that the SVM predictions of CEEMDAN perform better in 5 days' lag predictions for short-term and mid-term series and 15 days lag predictions for the long term series. From MAPE, we can also observe that the prediction of long term series is more accurate than the short term and mid-term series across 5 days, 15 days and 30 days lag predictions. The increased error variation in short-term series can be attributed to the considerable effects of random factors in the short-term. The error of the combined prediction of 5 days' lag is smaller than the other lags.

Considering the lag effects of 5 days, 15 days and 30 days in predicting the Bitcoin price, 5-day lag period of CEEMDAN predictions are relatively more accurate than 15 days and 30 days' lag predictions. This implies that the prediction of Bitcoin prices is largely governed by the recent variations and is less affected by historical performance. This is also supported by the fact that short-term and mid-term series were relatively better predicted in 5 days' lag model. Specifically, the percentage reduction of MSE in mid-term series is drastically lower in 5 days lag as compared to the other two lag periods. However, the long term series was better explained by the 15 days lag model albeit with a minor reduction in MSE. From this, we infer that a 5 days lag model of CEEMDAN predicts the prices with reduced error. Figure 6 shows the graphs of the predicted and actual values for short, mid, long and combined series of Bitcoin prices for the 5 days' lag model of CEEMDAN.

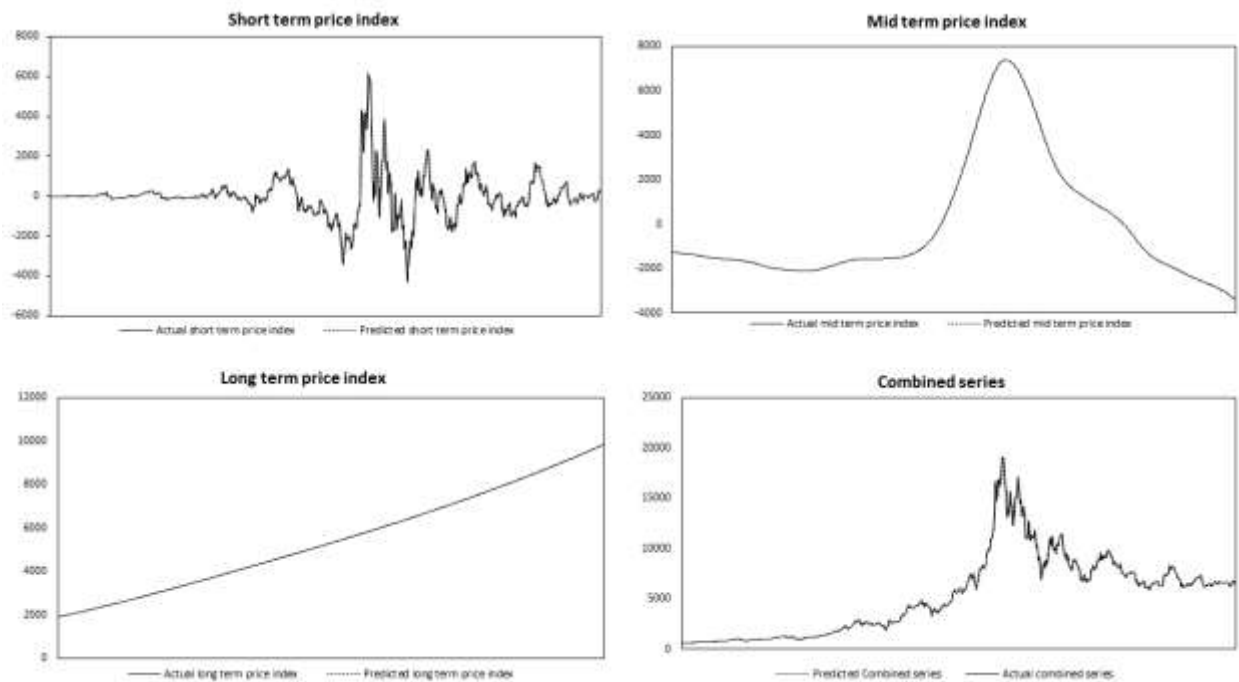


Figure 6 Comparison of predicted and actual values of the series for the 5 days' lag model

CONCLUSION

Cryptocurrencies have created a disruption in the financial markets and poses huge challenges for monetary policy makers. This also has raised questions to understand what qualities Bitcoin has as an asset. This study contributes to the emerging literature on Bitcoins in terms of understanding the dynamic nature of Bitcoin prices. It will enable to understand capabilities of Bitcoins in terms of its characteristics to be included in asset portfolios. It will also enable to reduce the ambiguity surrounding Bitcoins. The study began with an extensive literature review on the existing studies of Bitcoin in an efficient market. Based on previous research, the nature of Bitcoin as an asset is being contemplated (Baur, Dimpfl and Kuck, 2018). Therefore, the multiscale nature of the Bitcoin price series was examined using CEEMDAN analysis. The Bitcoin price series was decomposed into 10 intrinsic mode functions and reorganized with respect to its time-frequency reflecting the changes in Bitcoin price. The frequency was aggregated into three indices: the short-term price index, the medium-term price index and the long-term price index. They take 1.12%, 26.84%, 69.97% of the total variability of the observed Bitcoin prices respectively. The SVM-based prediction model is adopted for each price index to forecast results using 5 days, 15 days and 30 days lagged values. From the results/error metrics, it can be found that the short-term price index, consistent with the reaction of Bitcoin price, is speculative and affected by economic events. The fluctuation cycle of Bitcoins is observed to be of 5-days term, implying the self-adaptive nature of Bitcoin price in the short run. The long-term price index of Bitcoin grows steadily because it's an open but limited resource. The heuristics turn interpretable, as against EMH, if the short-term price index and the long-term price index were associated with Bitcoin's monetary property. The medium-term pattern supports EMH as it is not possible to consistently predict the Bitcoin price. However, the observations with respect to medium-term cannot be held with conviction as Bitcoin is recent and the analysis spans over a limited time-period. In future, with an extended time period, it may be possible to predict the medium-term price index as observed in the case of gold (Wen et al., 2017).

Thus, we achieve the intended objective to study the pattern of Bitcoin prices. The application of CEEMDAN sheds light on the distinct nature and price change of Bitcoin. The study, however, suffers from two limitations. Firstly, the study is limited to only a single cryptocurrency – Bitcoin. Other cryptocurrencies also need to be studied to identify their dynamic nature. Secondly, the study has only used an SVM based model for prediction of Bitcoin price series. Other machine learning algorithms can also be experimented to check the robustness of results, or maybe a comparative evaluation of different algorithms can give new insights.

REFERENCES

- Arasteh, A., Moradi, M. H., & Janghorbani, A. (2016). A novel method based on empirical mode decomposition for P300-based detection of deception. *IEEE Transactions on Information Forensics and Security*, 11(11), 2584-2593.
- Bariviera, A. F., Basgall, M. J., Hasperue, W., & Naiouf, M. (2017). Some stylized facts of the Bitcoin market. *Physica A*, 484(2), 82-90.
- Bartos, J. (2015). Does Bitcoin follow the hypothesis of efficient market? *International Journal of Economic Sciences*, 4(2), 100-114.
- Baur, D. G., Dimpfl, T., & Kuck, K. (2018). Bitcoin, gold and the US dollar—A replication and extension. *Finance Research Letters*, 25, 103-110.
- Bouoiyour, J., Selmi, R., Tiwari, A. K., & Olayeni, O. R. (2016). What drives Bitcoin price. *Economics Bulletin*, 36(2), 843-850.
- Bouri, E., Azzi, G., & Dyhrberg, A. H. (2017). On the return-volatility relationship in the Bitcoin market around the price crash of 2013. *Economics*, 11(2), 1.
- Bouri, E., Molnar, P., Azzi, G., Roubaud, D., & Hagfors, L. I. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? *Finance Research Letters*, 20(1), 192-198.
- Briere, M., Oosterlinck, K., & Szafarz, A. (2015). Virtual currency, tangible return: Portfolio diversification with Bitcoin. *Journal of Asset Management*, 16(6), 365-373.
- Brock, W. S. (1996). A test for independence based on the correlation dimension. *Econometric reviews*, 15(3), pp.197-235.
- Campbell, J. Y., & Perron, P. (1991). Pitfalls and opportunities: what macroeconomists should know about unit roots. *NBER macroeconomics annual*, 6, 141-201.
- Caporale, G. M., & Plastun, A. (2018). The day of the week effect in the cryptocurrency market. *Finance Research Letters*.
- Caporale, G., Gil-Alaña, L., & Plastun, A. (2018). Persistence in the cryptocurrency market. *Research in International Business and Finance*, 46(2), 141-148.
- Chang, C.-C., & Lin, C. J. (2011, 2: 27: 1--27: 27). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3), 27.
- Ciaian, P., Rajcaniova, M., & Kancs, D. A. (2016). The economics of Bitcoin price formation. *Applied Economics*, 48(19), 1799-1815.
- Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, 62, 182-199.
- Corbet, S., Meegan, A., Larkin, C., Lucey, B., & Yarovaya, L. (2018). Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, 165(1), 28-34.
- Drost, F. C., & Nijman, T. E. (1993). Temporal aggregation of GARCH processes. *Econometrica: Journal of the Econometric Society*, 909-927.
- Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar –A GARCH volatility analysis. *Finance Research Letters*, 16(2), 85-92.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2), 383-417.
- Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D., & Giaglis, M. G. (2015). Using time-series and sentiment analysis to detect the determinants of Bitcoin prices. *Information systems in a changing economy and society*, 371.
- Greaves, A., & Au, B. (2015). Using the Bitcoin transaction graph to predict the price of Bitcoin. Retrieved from www.snap.stanford.edu on 20th Decemeber, 2018
- Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and upto the minute prices. *The Journal of Finance and Data Science*, 4(2), 183-201.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., . . . Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series

analysis. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, (vol. 454, pp. 903-995).

Jang, H., & Lee, J. (2018). An empirical study on modeling and prediction of Bitcoin prices with bayesian neural networks based on blockchain information. *IEEE Access*, 6, 5427-5437.

Joseph, N. L., & Mazouz, K. (2010). Testing for overreaction and return continuations in stock price index returns. *International Journal of Strategic Decision Sciences (IJSDS)*, 1(2), 93-112.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economic Letters*, 158, 3-6.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters*, 158, 3-6.

Khuntia, S., & Pattanayak, J. (2018). Adaptive market hypothesis and evolving predictability of Bitcoin. *Economic Letters*, 167, 26-28.

Kim, K.-j. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1-2), 307-319.

Kristoufek, L. (2018). On Bitcoin markets (in) efficiency and its evolution. *Physica A: Statistical Mechanics and its Applications*, 503, 257-262.

Lahmiri, S. (2013). Forecasting direction of the S&P500 movement using wavelet transform and support vector machines. *International Journal of Strategic Decision Sciences (IJSDS)*, 4(1), 79-89.

Lim, K.-P., & Brooks, R. (2011). The evolution of stock market efficiency over time: a survey of the empirical literature. *Journal of Economic Surveys*, 25(1), 69-108.

McNally, S., Roche, J., & Caton, S. (2018, March). Predicting the price of Bitcoin using Machine Learning. In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)* (pp. 339-343). IEEE.

Ming, L., Yang, S., & Cheng, C. (2016). The double nature of the price of gold—A quantitative analysis based on Ensemble Empirical Mode Decomposition. *Resources Policy*, 47, 125-131.

Nadarajah, S., & Chu, J. (2017). On the inefficiency of Bitcoin. *Economic Letters*, 150(7), 6-9.

Nardo, M., Petracco-Giudici, M., & Naltsidis, M. (2016). Walking down wall street with a tablet: A survey of stock market predictions using the web. *Journal of Economic Surveys*, 30(2), 356-369.

Nawroth, A. P., & Peinke, J. (2006). Multiscale reconstruction of time series. *Physics Letters A*, 360(2), 234-237.

Nelson, D. M., Pereira, A. C., & de Oliveira, R. A. (2017, May). Stock market's price movement prediction with LSTM neural networks. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 1419-1426). IEEE.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259-268.

Peng, Y., Albuquerque, P. H. M., de Sá, J. M. C., Padula, A. J. A., & Montenegro, M. R. (2018). The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression. *Expert Systems with Applications*, 97, 177-192.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: Journal of the Econometric Society*, pp.1361-1401.

Ramirez, A. J., Rodriguez, E., & Valdez, I. C. (2018). Long-range correlations and asymmetry in the Bitcoin market. *Physica A*, 492(1), 948-955.

Shik Lee, H. (2004). International transmission of stock market movements: a wavelet analysis. *Applied Economics Letters*, 11(3), 197-201.

Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29(4), 309-317.

Torres, M. E., Colominas, M. A., Schlotthauer, G., & Flandrin, P. (2011). A complete ensemble empirical mode decomposition with adaptive noise. *Acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference*, 4144-4147.

- Tripathy, N. (2018). Predicting Stock Market Price Using Neural Network Model. *International Journal of Strategic Decision Sciences (IJSDS)*, 9(3), 84-94.
- Urquhart, A. (2016). The inefficiency of Bitcoin. *Economic Letters*, 148(1), 80-82.
- Wen, F., Yang, X., Gong, X., & Lai, K. K. (2017). Multi-Scale Volatility Feature Analysis and Prediction of Gold Prices. *International Journal of Information Technology & Decision Making*, 16(1), 205-223.
- Yonghong, J., He, N., & Weihua, R. (2017). Time-varying long-term memory in Bitcoin market. *Finance Research Letters*, 115(2), 1-5.
- Zargar, F. N., & Kumar, D. (2019). Long range dependence in the Bitcoin market: A study based on high-frequency data. *Physica A: Statistical Mechanics and its Applications*, 515, 625-640.
- Zhang, W., Wang, P., Li, X., & Shen, D. (2018). The inefficiency of cryptocurrency and its cross-correlation with Dow Jones Industrial Average. *Physica A*, 510, 658-670.
- Zhu, Y., Dickinson, D., & Li, J. (2017). Analysis on the influence factors of Bitcoin's price based on VEC model. *Financial Innovation*, 3(1), 3-16.
- Zivot, E. A. (1992). Further evidence on the great crash, the oil-price shock, and the unit root hypothesis. *Journal of Business and Economic Statistics*, 10, 251-270.