EXOGENOUS FACTORS THAT INFLUENCE RETURNS OF BITCOIN AND ETHER

Tomas Pečiulis – Asta Vasiliauskaitė

Abstract

Purpose of the paper - to determine if essential global economic factors are influencing returns of Bitcoin and/or Ether and if these cryptocurrencies react to forementioned economic factors as a phenomenon of one financial entity. In this paper we apply SARMAX family models to forecast the daily returns of the two main cryptocurrencies with the highest capitalization - Bitcoin and Ether the circulating cryptocurrency in the ethereum blockchain network. We analysed 10 essential global economic factors (Crude oil, Gold, EUR/USD, 10-yr bond, Vix, CMC Crypto 200, FTSE 100, Nikkei 225, DAX 30, S&P 500) to determine which of them have the highest weight for the prediction of the returns of the Bitcoin and Ether as a supplement to their autoregressive nature and determined 5 most important exogenous factors to predict the returns of the analysed cryptocurrencies. We used machine learning algorithms to set the weights of the SARMAX models dependent variables as well as the form of the model itself.

Key words: Bitcoin, Ether, machine learning, SARMAX, exogenous factors

JEL Code: G17, C58, C55

Introduction

Blockchain technology with cryptocurrencies as a main attribute of the technology recently are gaining prominence in the modern economy. This new technology has paved the way for the emergence of a new decentralized financial system that is independent of central banks or any other system operator. Operations in the blockchain network are secure and virtually impossible to counterfeit, making this technology exceptionally attractive and growing extremely fast. According to https://coinmarketcap.com, there are currently 4,841 different cryptocurrencies with a total capitalization of $ 2 trillion. Due to its wide application and rapid growth, this modern technology is extensively researched and received a lot of attention (Li & Wang, 2017; Kristjanpoller & Minutolo, 2018)

Cryptocurrencies are of interest to investors, central banks and other institutions that form economic policy, so the analysis and forecasting of these currencies is becoming an increasingly important area of research for financial forecasting theorists and practitioners who are looking for a new, innovative solutions, making the emergence of new forecasting methodologies progressively important issue (Kristjanpoller R & Hernández P, 2017; Monfared & Enke, 2014). Cryptocurrencies time series are highly volatile when compared to traditional currencies, which makes it a challenge to predict and a great opportunity to implement machine learning techniques (Kristjanpoller & Minutolo, 2018).

The vast majority of the studies concentrate on Bitcoin as a benchmark of blockchain technology. However, this technology is rich in other more versatile applications, such as ether, so this study will examine the cryptocurrency ether used in the ethereum crypto network in tandem with bitcoin. The unique set of economic factors were used to supplement auto-correlation models to determine whether these two currencies react to forementioned economic factors as a phenomenon of one financial entity.

1 Literature review

The literature analysis is focused on two aspects. Firstly, the scientific literature is analysed on machine learning in financial forecasting. Secondly, literature on cryptocurrencies. Those two aspects are chosen because cryptocurrencies time series are highly volatile when compared to traditional currencies, which makes it a challenge to predict and a great opportunity to implement machine learning techniques (Kristjanpoller & Minutolo, 2018) and cryptocurrencies are extensively researched and currently received a lot of attention in the scientific world (Li & Wang, 2017).

Measuring, forecasting and modelling the value of different financial instruments, such as commodities, exchange rates, stock prices, etc., is a widely discussed topic in financial research. Even a small improvement in models that forecasts returns or prices of financial assets would lead to a better understanding of the market and better decisions not only for central banks and economic policy makers, but also for private investors, making application of machine learning in financial forecasting an increasingly explored field in the scientific world (Kristjanpoller & Minutolo, 2018).

Machine learning algorithms have been used for many years to obtain financial forecasts. (White & Diego, 1988) were among the first to show that the use of artificial intelligence methods can provide statistically significant results in predicting the returns of IBM stock. (Gu et al., 2020) studied various machine learning methods for pricing the S&P 500, and their results revealed that using a machine learning element can yield several times more accurate results compared to simple regression methods. (Kaczmarek & Perez, 2021) used machine learning to forecast the value of securities and, based on the results obtained, formed investment portfolios. Using this selection method, the authors obtained an 18% better result than they would have used the equal weight securities selection method.

Thus, it can be seen that financial forecasting is a widely discussed scientific topic and the element of machine learning increasingly used in various studies often yields more accurate forecasting results, so it makes machine learning one of most promising forecasting tools.

Cryptocurrencies are extensively analysed in the scientific literature, but there is still a lack of research compared to other fields (Pichl & Kaizoji, 2017). Various disciplines analyse this phenomenon - in law it is analysed to understand what regulatory mechanisms and legislation should be adopted to ensure optimal functioning of cryptocurrencies (LAN JU, 2016; Böhme et al., 2015); capital markets, to assess whether bitcoin as a major cryptocurrency can be classified as a currency system of an emerging market (Carrick, 2016); econometrics, in an attempt to determine what economic factors affect the price of bitcoin (Li & Wang, 2017); environmental, to explore how blockchain technology could contribute to more sustainable economic growth (Upadhyay et al., 2021).

The main advantage of cryptocurrencies is that they enable decentralized financial, economic and even political systems (Estevam et al., 2021). The main disadvantage of cryptocurrencies is extreme volatility of their prices (Li & Wang, 2017), so the analysis and forecasting of these currencies is becoming an increasingly important area of research (Kristjanpoller R & Hernández P, 2017).

2 Methodology

This section describes the data used to perform the study, their processing, and the study procedures. SARMAX (seasonal autoregressive moving average with exogenic variable) family models were used to conduct whether ETH-USD and BTC-USD react to researched economic factors as a phenomenon of one financial entity. SARMAX family models were chosen because they include autoregressive (AR), moving average (MA) and seasonal (S) aspects as well as they accept exogenous variables (X). SARMAX(p, q)(m, P, Q) takes 5 inputs that have to be optimized, p and q are number of lags of AR and MA parts of the model respectively, P,Q refer to the autoregressive, differencing, and moving average terms for the seasonal part of the model, m refers to the number of periods in each season.

The study was conducted using ETH-USD and BTC-USD daily exchange rates. Data obtained from https://finance.yahoo.com/. As a larger amount of data guarantees more accurate statistical results, the maximum period during which the data were published was chosen, in this case since the first day the ether and the dollar exchange rate were first officially published - 08/07/2015 to 24/04/2021. Daily closing prices were used for further analysis. We analysed 10 essential global economic factors to determine which of them have the highest weight for the prediction of the returns of the Bitcoin and Ether as a supplement to their autoregressive nature. These global economic factors are:

- ***Gold*** and ***Crude oil*** represents commodity market and was used to test how this market influences our researched topic
- ***EUR/USD*** is the biggest market of classical currencies, so it is necessary to test to what extent it influences crypto market.
- ***10-yr US treasury bonds.*** These bonds are the main instrument of the monetary policy of the US, the biggest economy in the world, so it is necessary to test to what extent monetary policy of the US influences crypto market.
- ***VIX index.*** The VIX Index reflects traders 'expectations for the next 30 days of stock market risk, calculated based on short-term option prices in the S&P 500. As the expectation is one the demand and(or) supply curve shifters, so it is necessary to test to what extent expectations on stock market prices influences crypto market.
- ***CMC Crypto 200 index.*** The Index intends to track the price movements of a portfolio of the top 200 cryptocurrencies by market capitalization, so it is necessary to test to what extent the whole crypto market influences fluctuation, of ETH/USD and BTC/USD.
- ***FTSE 100 index, Nikkei 225 index, DAX 30 index, S&P 500 index.*** 4 most influential stock indexes from different continents or different countries. Represents how the fluctuation of the stock market influences crypto market

As the SARMAX family models are not applicable to the analysis of non-stationary data, it was necessary to investigate the stationarity of the obtained data, to determine the seasonality, and to check whether these data were distributed according to the normal distribution. The Q-Q plot (quantile-quantile plot) shown in the Fig. 1 was used for the initial determination of normality. The Q-Q plot (quantile-quantile plot) was used for the initial determination of normality. A seasonal decomposition graph was used to determine seasonality. An augmented Dickey-Fuller test was used to determine the stationarity of the data. Prices were transformed to returns using 1 formula. There r_t is return of today, P_t price of today, p_{t-1} price of yesterday.

$$r_t = \frac{P_t}{p_{t-1}} * 100 \qquad (1)$$

The suitability of the models was assessed in two aspects. First, the formula of the model should not have statistically insignificant coefficients, second, the akaike information criterion (AIC) was evaluated, for which formula 2 was used.

$$AIC = 2k - 2ln(L), \qquad (2)$$

In this formula k is the number of the parameters and L is the result of maximum likelihood estimator. It can be seen from the AIC formula that the value of this criterion decreases when maximum likelihood estimator increases, but it increases whenever a new parameter is added, so we are looking for a model with a lowest value of AIC.

The parameters of the models were obtained with the help of machine learning using predefined functions in python open-source program, "pmdarima" library[1]. The data was split into the two sets. Training set consisted of 80 % of the data and the testing set consisted of 20% of the data. The parametrization was performed fully automatically, it means that architecture of the model was found using machine learning, including determination of the optimal number of lags and finding optimal coefficients for the parameters. Automatic parametrization drastically reduces time used for model optimization. Moreover, it ensures that optimal architecture of the model will be achieved.

3 Empirical research and results

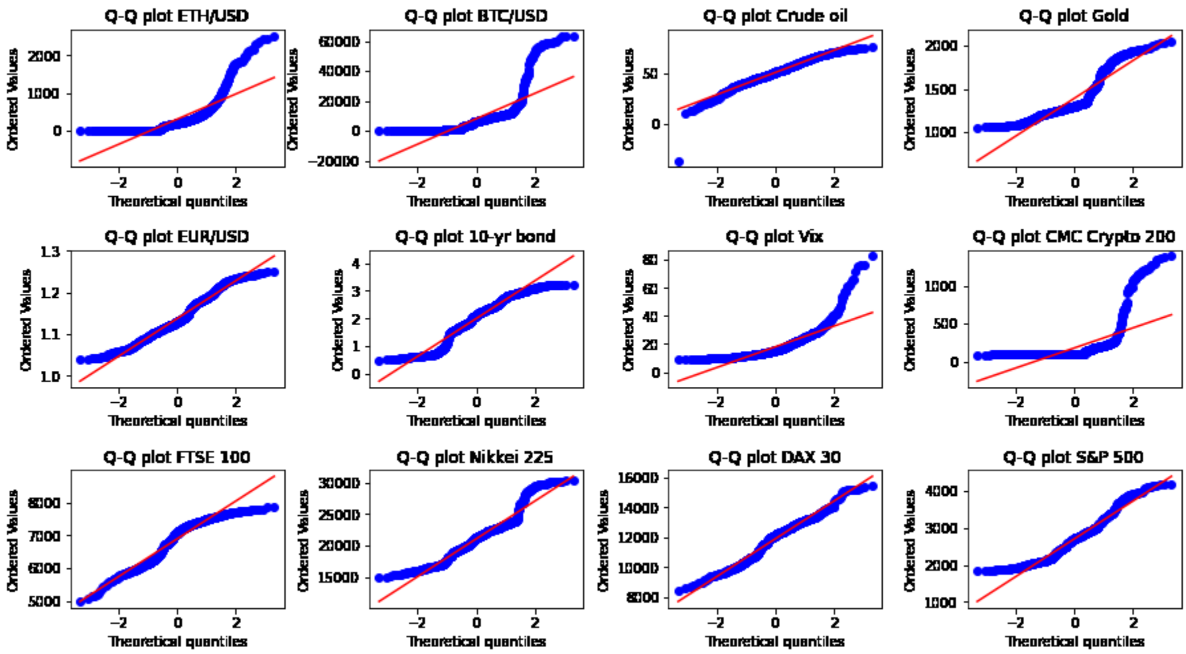
As the SARMAX family models are not applicable to the analysis of non-stationary data, it was necessary to investigate the stationarity of the obtained data, to determine the aspect of seasonality, and to check whether these data were distributed according to the normal distribution. The Q-Q plot (quantile-quantile plot) shown in the Fig. 1 was used for the initial determination of normality.

It is clear from the graph provided that the data are not distributed according to the normal distribution. Because the results of the Q-Q graph were so unambiguous, there was no need to perform any additional test that would mathematically assess the normality of the distribution.

A seasonal decomposition graph was used to determine seasonality. It showed that none of the time series showed a seasonal aspect.

[1] Auto-arima function was use for the optimization of the parameters

Fig. 1: Q-Q plots



Source: made by authors

An augmented Dickey-Fuller test was used to determine the stationarity of the data. The results are shown in the Tab. 1

Tab. 1: Augmented Dickey Fuller test results of the prices

	ETH/ USD	BTC/ USD	Crude oil	Gold	EUR/U SD	10-yr bond	Vix	CMC Crypto 200	FTSE 100	Nikkei 225	DAX 30	S&P 500
ADF Statistic	1.79	1.69	-2.24	-0.52	-2.09	-1.08	-4.09	4.60	2.47	-0.80	-1.85	0.23
p-value	0.99	1.00	0.19	0.89	0.25	0.72	0.01	1.00	0.12	0.82	0.35	0.97
Critical 1 %	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44
Critical 5 %	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86
Critical 10 %	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57

source: made by authors

As we can see from the table almost all prices are not stationary, because ADF statistics values are bigger than critical value (even with a10% level of significant) with the exception of the wix index (-4.09), but to maintain solidity of the data all prices were transformed to returns using 1 formula.

After the transformation augmented Dickey Fuller test was performed once more. The results are shown in the Tab. 2.

Tab. 2: Augmented Dickey Fuller test results of the returns

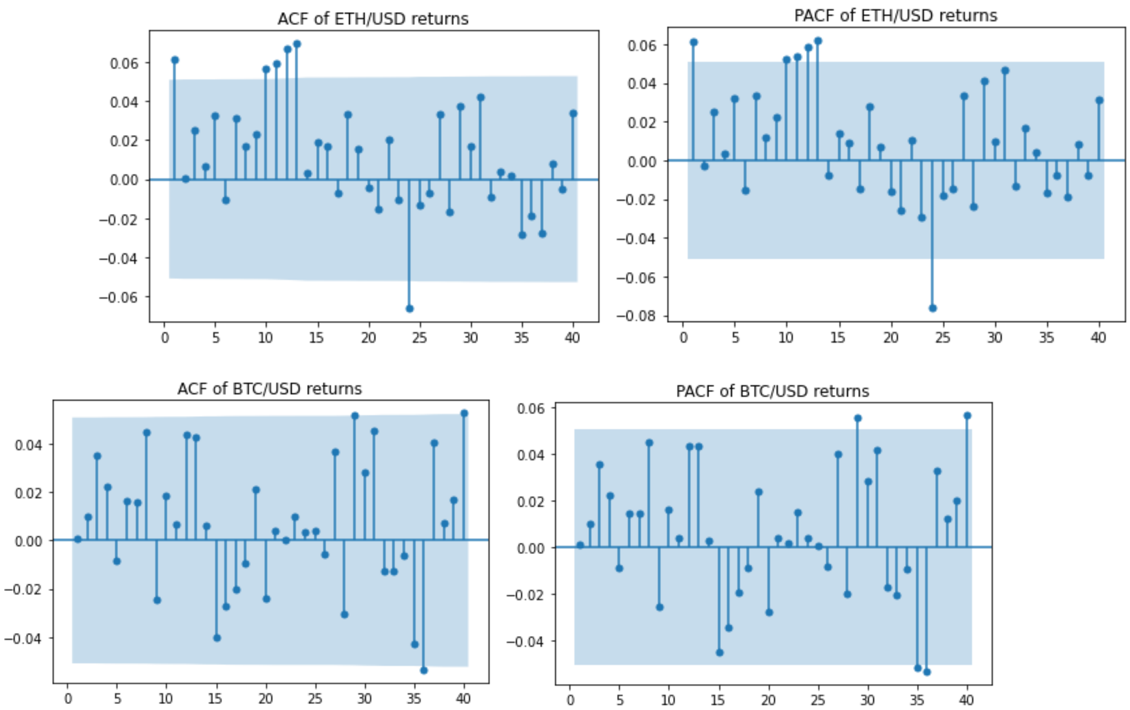
	ETH/ USD	BTC/U SD	Crude oil	Gold	EUR/U SD	10-yr bond	Vix	CMC Crypto 200	FTSE 100	Nikkei 225	DAX 30	S&P 500
ADF Statistic	-8.08	-38.51	-8.36	-17.49	-38.06	-14.57	-17.11	-37.89	-12.45	-26.26	-12.57	-11.03
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Critical 1 %	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44
Critical 5 %	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86	-2.86
Critical 10 %	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57

source: made by authors

As we can see from the table all returns are stationary, because ADF statistics values are smaller than critical values, so further analysis was performed with transformed data.

The parameters of the models were obtained with the help of machine learning using a predefined functions in python open-source program as it is described in methodology section. However, to protect models from overfitting we used autocorrelation function (ACF) and partial autocorrelation function (PACF) to find out the optimal p and q values. We generated ACF and PACF graphs for the dependent variables (ETH/USD and BTC/USD). The ACF and PACF graphs are shown in Fig. 2.

Fig. 1: ACF and PACF for ETH/USD and BTC/USD



source: made by authors

From Fig. 2 we can see that neither of the cryptocurrencies returns demonstrates strong autoregressive or moving average relationship, so we do not expect to get a model with high p or q values. Moreover, it is great proof that models without exogenous factors will show weak performance in forecasting BTC/USD or ETH/USD exchange rates and it returns, so this confirms even bigger relevance of exogenous variables to this research.

Tab. 2: Augmented Dickey Fuller test results of the returns

	Crude oil	Gold	EUR/ USD	10-yr bond	Vix	CMC Crypto 200	FTSE 100	Nikkei 225	DAX 30	S&P 500
BTC/ USD	875.00	873.00	875.00	875.00	873.00	843.00	873.00	875.00	873.00	871.00
p-value	0.12	0.00	0.78	0.20	0.00	0.00	0.00	0.69	0.00	0.00
ETH/ USD	1038.00	1037.00	1038.00	1038.00	1037.00	1027.00	1037.00	1038.00	1037.00	1036.00
p-value	0.84	0.00	0.61	0.56	0.00	0.00	0.00	0.76	0.00	0.00

source: made by authors

It can be seen from the data in the table that BTC/USD returns have a higher relationship with exogenous variables in general. Moreover, Crude oil, EUR/USD exchange rate, 10-yr US treasury bonds, Nikkei 225 has insignificant p-values, so these exogenous variables do not provide any additional explanation to the movement of the returns, so they should not be included in the analysis of neither BTC/USD nor ETH/USD. On the other hand, gold, Vix index, CMC Crypto 200, FTSE 100, DAX 30, S&P 500 had p-values equal to zero and lower AIC values, so they can be included in BTC/USD and ETH/USD, but a further investigation is necessary to check for perfect multicollinearity, e. g. FTSE 100 and DAX 30 may be representing the strength of Europe's economy and including both of them would result in unnecessary exogenous variables. To sum up, both BTC/USD and ETH/USD reacts to the exogenous variables identically, so it safe to say that these cryptocurrencies are phenomenon of one financial entity.

The results are somehow surprising and raises further questions. Why do some commodities (e.g., gold) have high statistical value for explaining movement of cryptocurrencies and some of them (e.g., crude oil) are insignificant? Why do some stock indexes significant and some are not? Finally, further research for the best combination of economic factors in the VAR model is needed.

Conclusion

1 Investigation of the recent scientific literature has shown that the analysis and the prediction of cryptocurrencies movement is a widely discussed scientific topic. The increasingly popular element of machine learning often yields more accurate prediction results. The advantages and disadvantages of cryptocurrencies are also revealed. The main shortcoming of cryptocurrencies has been identified as their value instability, so developing more sophisticated methods provide significant benefits to investors and economic decision makers.

2 The initial analysis of the data revealed that the time series data were not distributed according to the normal distribution, so statistical methods that require this condition are not suitable for their analysis. Also, no seasonality element was detected after the formation of the seasonal decomposition graph, so seasonal models are not required to analyse the data. Finally, an extended Dickey-Fuller test was performed, which showed that the data were not stationary, so a data transformation was performed and returns of the prices were formed, with which the same test was performed again and the results showed that the data were stationary.

3 BTC/USD returns have a higher relationship with exogenous variables in general. Moreover, Crude oil, EUR/USD exchange rate, 10-yr US treasury bonds, Nikkei 225 has insignificant p-values, so these exogenous variables do not provide any additional explanation to the movement of the returns of any of analysed cryptocurrencies, so they should not be included in the analysis of neither BTC/USD nor ETH/USD. On the other hand, gold, Vix index, CMC Crypto 200, FTSE 100, DAX 30, S&P 500 had p-values equal to zero and lower AIC values, so they can be included in BTC/USD and ETH/USD analysis and forecasting, but a further investigation is necessary to check for perfect multicollinearity. Finally, both BTC/USD and ETH/USD reacts to the exogenous variables identically, so it safe to say that these cryptocurrencies are phenomenon of one financial entity.

References

Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, Technology, and Governance. *Journal of Economic Perspectives*, *29*(2), 213–238.

Carrick, J. (2016). Bitcoin as a Complement to Emerging Market Currencies. *Emerging Markets Finance and Trade*, *52*(10), 2321–2334. https://doi.org/10.1080/1540496X.2016.1193002

Estevam, G., Palma, L. M., Silva, L. R., Martina, J. E., & Vigil, M. (2021). Accurate and decentralized timestamping using smart contracts on the Ethereum blockchain.

Information Processing and Management, *58*(3), 102471. https://doi.org/10.1016/j.ipm.2020.102471

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Kaczmarek, T., & Perez, K. (2021). Building portfolios based on machine learning predictions. *Economic Research-Ekonomska Istraživanja*, *0*(0), 1–19. https://doi.org/10.1080/1331677x.2021.1875865Kristjanpoller R, W., & Hernández P, E. (2017).

LAN JU, T. (JUN) L. A. Z. T. (2016). Capital Flight and Bitcoin Regulation. *International Review of Finance*, *16*, 445–455.

Li, X., & Wang, C. A. (2017). The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. *Decision Support Systems*, *95*, 49–60. https://doi.org/10.1016/j.dss.2016.12.001

Monfared, S. A., & Enke, D. (2014). Volatility forecasting using a hybrid GJR-GARCH neural network model. *Procedia Computer Science*, *36*(C), 246–253. https://doi.org/10.1016/j.procs.2014.09.087

Pichl, L., & Kaizoji, T. (2017). Volatility Analysis of Bitcoin Price Time Series. *Quantitative Finance and Economics*, *1*(4), 474–485. https://doi.org/10.3934/qfe.2017.4.474

Upadhyay, A., Mukhuty, S., Kumar, V., & Kazancoglu, Y. (2021). Blockchain technology and the circular economy: Implications for sustainability and social responsibility [Article]. *Journal of Cleaner Production*, *293*. https://doi.org/10.1016/j.jclepro.2021.126130

White, H., & Diego, S. (1988). Economic prediction. *Proceedings of The IEEE International Conference on Neural Networks*.

Contact

Tomas Pečiulis

Vilniaus Gedimino technikos universitetas (VILNIUS TECH)

Saulėtekio al. 11, Vilnius 10223

tomas.peciulis@vilniustech.lt

Asta Vasiliauskaitė

Vilniaus Gedimino technikos universitetas (VILNIUS TECH)

Saulėtekio al. 11, Vilnius 10223

asta.vasiliauskaite@vilniustech.lt