

Forecasting cryptocurrencies prices

Kamil Matuszelański, 387078

Table of Contents

Abstract	1
Introduction	2
Dataset description.....	2
Analysis.....	4
Bitcoin price forecasting	4
Ethereum price forecasting.....	8
Forecasting both series using VAR model	10
Models comparison.....	11
Conclusions	13
References	13

Abstract

This study is aimed at price prediction of two most popular cryptocurrencies - Bitcoin and Ethereum. I have used 2 popular approaches to time series analysis, namely ARIMA and Vector AutoRegressive models. One of the goals of this study was comparison of performance of these two methods, directly measured by performance on out-of-sample period. I have shown that ARIMA models generally outperform VAR approach in terms of goodness of fit for out-of-sample period.

Introduction

Nowadays cryptocurrency market is growing rapidly. Current value of the whole sector is estimated at 1.03 Billion dollars, and it is expected to rise to 1.4 billion by 2024. Investors are interested in this market because of independence of classical investment commodities and the premise of astonishing returns. For example, 2017 only, value of the first and the most popular currency, Bitcoin (BTC), rose by 1318%. Although this was a clear bubble and the return next year was -72%, the interest of larger audiences in the whole marked stayed at a high level.

This study is devoted to price prediction of two of the most popular cryptocurrencies, Bitcoin and Ethereum. The results could help investors make better decisions concerning asset allocation into these commodities. Secondary goal of this study was to compare results of using Vector Auto Regressive model and casual ARIMA models on the two series. Because the series are cointegrated only over shorter time period, I have created ARIMA models for both long, 1.5 year period, and shorter one, spanning a little over 1 month. This way results of ARIMA and VAR can be more fairly compared.

Contents of this paper is as follows. First section deals with basic dataset analysis, including analysis of integration order and assessment of cointegration of two series. Second section is a description of modeling process, first using ARIMA models, and then estimating VAR model. Finally, comparison of created models and summary of results are presented.

Dataset description

I have started my analysis with obtaining basic information about the two series. I have restricted my study to the period between 10.2018 and 04.2020. On plot 1, both series are presented. The differenced series is also included. From visual inspection, one can conclude that both of the series are integrated of order 1, and first difference is enough to make both series more or less stationary.

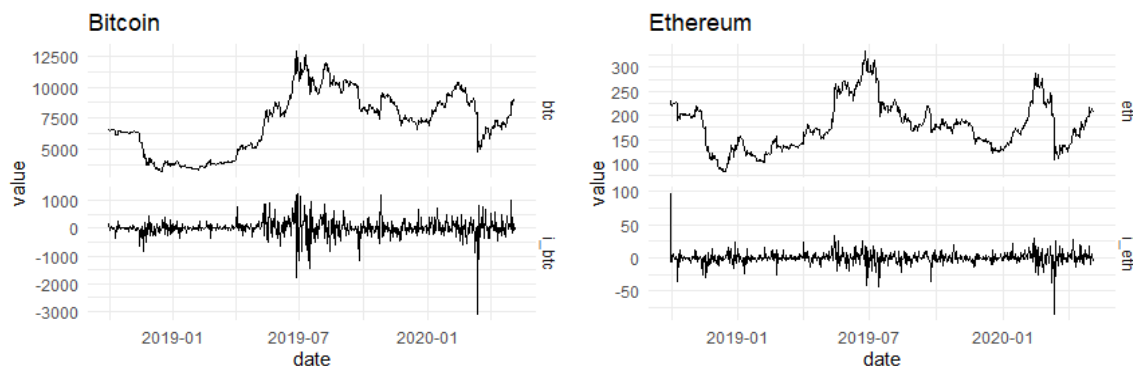


Figure 1: Bitcoin and Ethereum quotations, with no transformation and with differencing of order 1

The same result can be obtained from Augmented Dickey Fuller test for both variables. Test statistics are presented in table 1. From Breusch-Godfrey test for autocorrelation of residuals, one can conclude that correct number of augmentations to include in ADF for BTC is 1. With $p\text{-value} = 0.55$, one can conclude that the Bitcoin series is integrated of order at least 1. For differenced series, no augmentations are needed, and series is stationary ($p\text{-value} < 0.01$).

The same conclusions can be drawn from ADF test for Ethereum series - it is integrated of order 1. This could be expected, and is important for later analysis, as first requirement for both series to be cointegrated is fulfilled.

Table 1: Augmented Dickey-Fuller test for stationarity of Bitcoin and Ethereum series.

Series	augmentations	ADF statistic	ADF p-value	B-G test statistic value	B-G test p-value
BTC	0	-1.537581	0.4859518	7.2540578320	0.007074098
	1	-1.363869	0.5508909	0.0054504387	0.941147974
diff(BTC)	0	-26.314128	0.0100000	0.0044221659	0.946980261
ETH	0	-2.230838	0.2267899	3.1410305665	0.076345554
diff(ETH)	0	-25.413805	0.0100000	0.0002061795	0.988543609

I have also tested second condition for cointegration of two time series, specifically stationarity of residuals from the linear model, as the second step of the Engle-Granger method. On the plot 2 residuals from regression Bitcoin over Ethereum are presented. From visual inspection one can conclude that the residuals series is non-stationary. This is also confirmed by ADF test (p.value = 0.397). This leads to conclusion that series are not cointegrated on their full time span, and applying VAR model is impossible.



Figure 2: Residuals obtained from linear model in Engle-Granger method for both series

However, two series appear to be cointegrated on smaller time span, that is from 2020-01-31 to 2020-03-31. I have again tested this hypothesis by estimating linear model and indeed the residuals are stationary, with p.value < 0.01. From this, I have concluded that the series are cointegrated over that time period. On the plot 3, both series are presented. Values are max-scaled, as the price levels for BTC and ETH are very different (9731 vs 212 USD as of 09.05.2020, respectively).

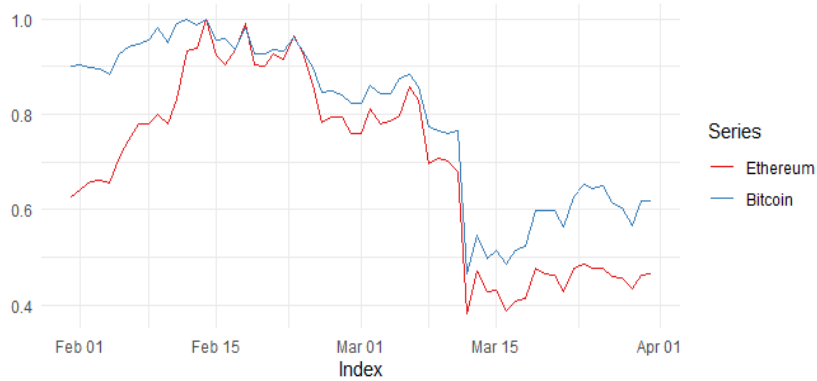


Figure 3: BTC and ETH series for smaller period, max-scaled.

On the below plot 4, I have visualized the periods used for later analyses. Long period, indicated by color red and green (2018-10-01 to 2020-03-31), was used to estimate ARIMA model, while short period (green color only, 2020-01-31 to 2020-03-31) was used to estimate both ARIMA and VAR models. Out-of-sample period from 2020-04-01 to 2020-05-05 is presented with blue. Observations from this period were used to validate the models.

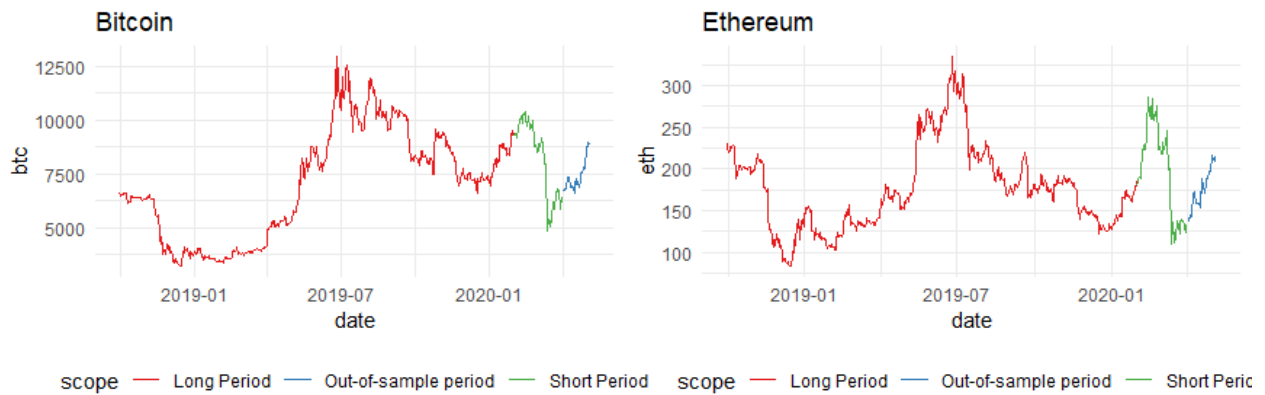


Figure 4: Split of both series for long, short and out-of-sample periods.

Analysis

Bitcoin price forecasting

As the first one I have estimated model for long period for Bitcoin price. Long period accounts for 548 observations. As stated before, the series is $I(1)$, so correct model will be $ARIMA(p, 1, q)$.

To determine a possible order of the model I have created ACF and PACF plots, shown on plot 5. For ACF, most significant lags are 1, 17 (mildly), 18 and 20. For PACF, these are 1, 9 and 20. One should note that such long period of lag could lead to an overly variable-inflated model. This can cause practical problems. Such specification should be used only in very long series. For instance, estimating the same model but for smaller period of only 60 observations would be problematic. With maximum included lag 20, one third of the observations would not be used in estimation directly.

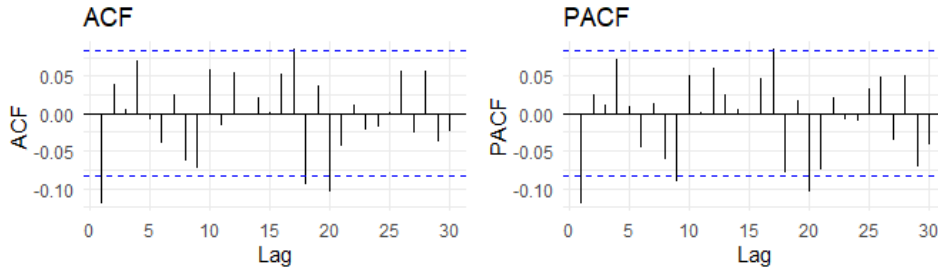


Figure 5: ACF and PACF for $\text{diff}(\text{BTC})$.

As the first baseline model, I have estimated the one with specification $\text{ARIMA}(1, 1, 1)$. Ljung-Box test with $p\text{-value}=0.17$ indicates no autocorrelation of residuals. I have created ACF and PACF plot 6 of residuals to assess autocorrelations of specific lags. Lag 20 seems to be still significant, while importance of other lags shown to be important before dropped below 5% significance threshold. In the table 2 I have included summary of the model. Z test of coefficients shows that both AR and MA parts seem to be insignificant ($p\text{-value}$ 0.43 and 0.64, respectively).

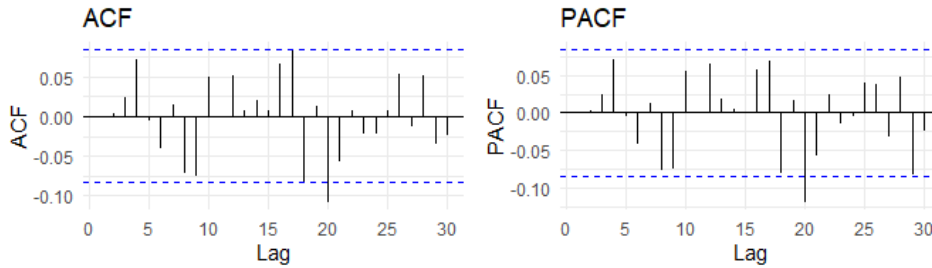
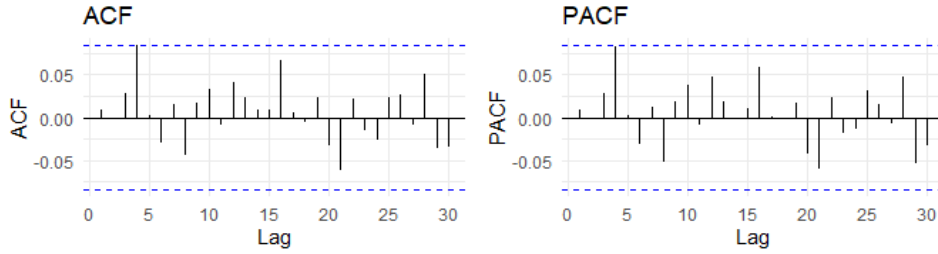


Figure 6: ACF and PACF of residuals for model $\text{ARIMA}(1,1,1)$ with BTC series.

Table 2: Summary of model $\text{ARIMA}(1,1,1)$.

term	estimate	std.error	statistic	p.value
ar1	-0.3008004	0.3834511	-0.7844559	0.4327727
ma1	0.1839386	0.3948397	0.4658563	0.6413184

To mitigate autocorrelation on lag 20 indicated by ACF plot, I have created a model $\text{ARIMA}(20,1,20)$, but only with lags: 1, 9, 17, 18 and 20. P-value for Ljung-Box test being essentially zero indicates that residuals are autocorrelated. However, analysis of autocorrelation plots shows the contrary. No correlation of lags seem to be significant. As of coefficients significance, only ma1 is significant, however some of the coefficients could not be computed by ARIMA optimizer, despite changing increasing the number of iterations.



I have also tried to fit ARIMA model automatically with *forecast* package. Using Akaike Information Criterion as goodness-of-fit measure, best model chosen was ARIMA(2, 1, 9). Within this model Ljung-Box test shows lack of autocorrelation of residuals (p-value = 0.09). On the plot 7 ACF and PACF functions are presented. Lag 20 is still significant. All coefficients for which standard errors could be computed seem to be insignificant, as shown in the table 3.

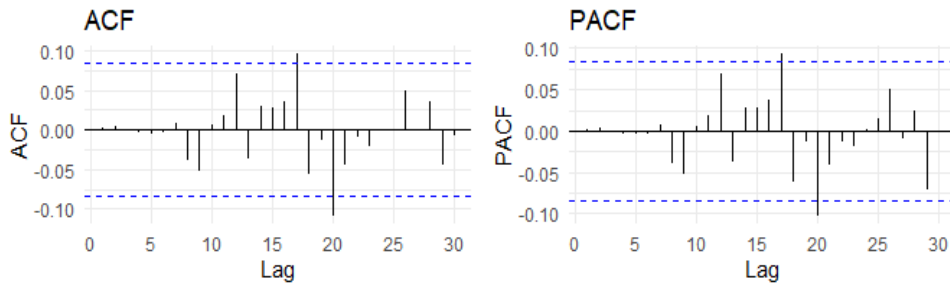


Figure 7: ACF for residuals of ARIMA(2,1,9) model

Table 3: Summary of model ARIMA(2,1,9).

term	estimate	std.error	statistic	p.value
ar1	-0.90377522	NaN	NaN	NaN
ar2	-0.92004046	NaN	NaN	NaN
ma1	0.79515565	NaN	NaN	NaN
ma2	0.86966874	NaN	NaN	NaN
ma3	-0.07080841	0.06580462	-1.0760402	0.28190927
ma4	0.13114233	0.06775885	1.9354273	0.05293788
ma5	0.08832036	0.07089832	1.2457327	0.21286253
ma6	0.03819121	0.06574193	0.5809263	0.56129015
ma7	-0.02493630	0.06498064	-0.3837497	0.70116397

term	estimate	std.error	statistic	p.value
ma8	-0.06049744	0.05784811	-1.0457981	0.29565426
ma9	-0.04315530	0.02934785	-1.4704759	0.14143291

According to the automatic selection, model ARIMA(0,1,1) was the best candidate using BIC criterion.

I have compared models' performance using AIC and BIC criterion. I have also used validation on out-of-sample period using RMSE measure. Test set consisted of 35 observations for period from 01.04.2020 to 05.05.2020. I have used one-step-ahead forecast, rather than estimation for the whole test period at once. There was number of reasons for that. First, forecasting from ARIMA models becomes very unreliable after longer period, as predictions regress to the mean value of the series. Second, estimating financial assets prices is a very non-trivial task, and variance of possible scenarios grows rapidly with increment in number of steps ahead.

In the below table 4, models introduced previously are presented. The table is sorted by RMSE error on out-of-sample data. As can be seen, best model by this criterion has included lags 1,9,17,18 and 20. On the other hand, by AIC measure the best is the same mode, but with removed lags 18 and 20. And finally, best model with BIC measure is ARIMA(0,1,1).

Table 4: Models comparison

Model	df	AIC	BIC	RMSE in sample	RMSE out of sample
Included lags 1, 9, 17, 18, 20	11	7883.073	7930.422	318.8903	257.4691
Included lags 1, 9, 17	7	7879.003	7909.134	320.0877	258.2616
ARIMA(2,1,9) (automatically, criterion AIC)	12	7889.300	7940.953	320.0715	263.6304
ARIMA(1,1,1)	3	7891.068	7903.981	326.2591	263.9751
ARIMA(0,1,1) (automatically, criterion BIC)	2	7889.869	7898.478	326.4991	265.1994

Next step of my analysis was again modeling Bitcoin price using ARIMA model, however this time I have used smaller number of observations to be able to compare the results to the ones from VAR model. Similarly as in long-period case, I have started with assessment of autocorrelation functions of differenced series, shown on the plot 8. Both figures indicate that there is virtually no autocorrelation of residuals, and as such, no place to actually use ARIMA

model. Only lag 4 in ACF function slightly crosses the edge of significance, and lag 1 is slightly below the significance threshold.

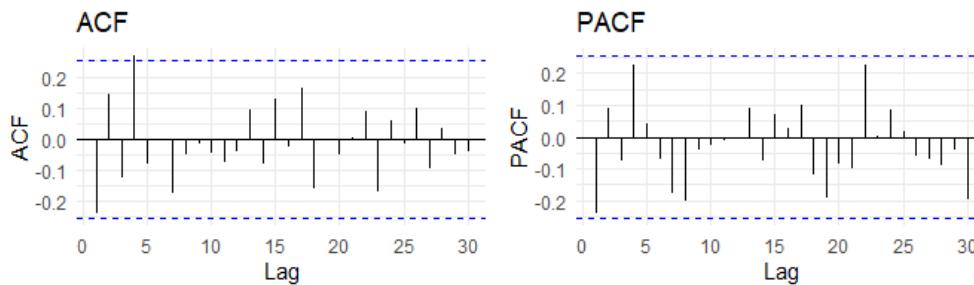


Figure 8: ACF and PACF function for BTC in short period

Nevertheless, I have tried fitting the following models:

- ARIMA(4,1,4)
- ARIMA with included lags ar4, ma4
- ARIMA(1,1,0), order selected automatically by AIC criterion.

Metrics of goodness of fit are presented in table 4. According to AIC, BIC and RMSE on test set, the best model is the most parsimonious one, ARIMA(1,1,0), or AR(1). The p-value for first lag in AR part of the model is equal to 0.07, meaning that at 5% significance level the parameter is insignificant and should be dropped. This again indicates that the task of predicting Bitcoin price using such short timespan is questionable.

Table 4: Comparison of goodness-of-fit metrics for BTC price in short period.

Model	df	AIC	BIC	RMSE in sample	RMSE out of sample
ARIMA(1,1,0) (automatically, criterion AIC)	2	919.9985	924.1872	495.5588	264.4412
Included lag 4	3	919.9313	926.2143	485.7367	286.1728
ARIMA(4,1,4)	9	923.5144	942.3635	436.8931	296.4631

Ethereum price forecasting

For Ethereum cryptocurrency, I have conducted similar analysis as for Bitcoin. First, I have estimated ARIMA model for long, 1.5 year period, and then for short, 2-month period in which the series is cointegrated with Ethereum.

Judging by ACF and PACF plots 9, only 4 and 20 lags seem to be significant, with lag 1 at the verge of significance.

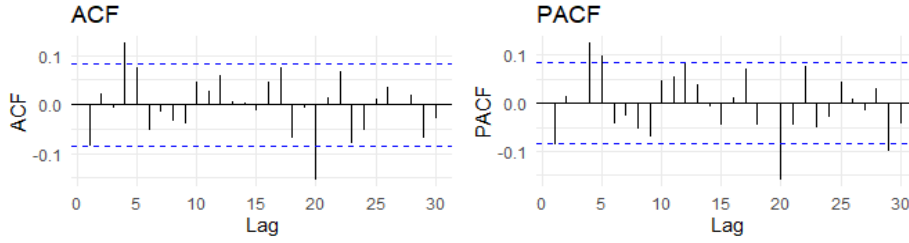


Figure 9: ACF and PACF function for ETH in long period

For clarity, I have not included full description of the steps taken during model estimation, as these were similar as in Bitcoin price modeling. The models I have consecutively tried were:

- ARIMA(1,1,1)
- ARIMA(20,1,20), with only with lags 4 and 20
- ARIMA(20,1,20) with all lags
- ARIMA(4,1,4) - model obtained using automatic order selection according to Akaike criterion
- ARIMA(4,1,4), with added lag 20.

I have created goodness-of-fit measures comparison for models estimated, shown in table table 5. The best model according to RMSE for out-of-sample period is ARIMA(20,1,20) with all lags. That model is also rated as second-best by AIC criterion. On the contrary, that the model is rated as the worst using BIC criterion. Model rated the highest by AIC criterion is ARIMA(4,1,4) model with added lag 20, while by BIC criterion - one with included lags 4 and 20.

Table 5: Comparison of goodness-of-fit measures for ARIMA models in ETH long series.

Model	df	AIC	BIC	RMSE in sample	RMSE out of sample
ARIMA(20,1,20)	41	3999.070	4175.552	8.553134	8.319748
ARIMA(4,1,4) (automatically, criterion AIC)	9	4000.123	4038.863	9.198562	8.534926
ARIMA(1,1,1)	3	4022.077	4034.990	9.498583	8.678306
ARIMA(4,1,4), with included lag 20	11	3989.141	4036.490	9.068901	8.899444
Included lags 4, 20	5	4003.596	4025.118	9.298356	9.051324

Similarly as for Bitcoin, also for Ethereum I have estimated model on shorter, 2-month series. On the plot 10 I have shown ACF and PACF functions. It is visible that the only significant lag is 4.

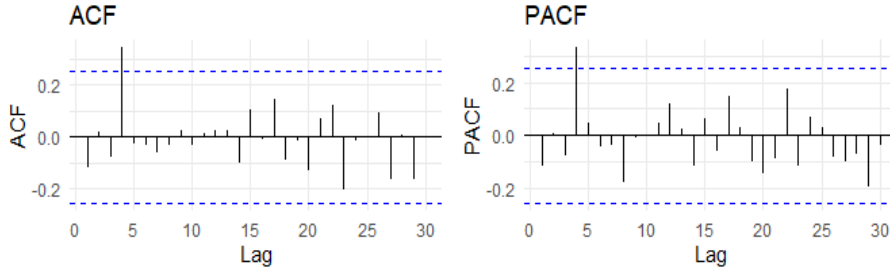


Figure 10: ACF and PACF function for ETH in short period

The models consecutively estimated by me are:

- ARIMA(4,1,4),
- Included lags AR(1,4) and MA(1,2,3,4),
- Included lag 4,
- ARIMA(0,1,4) (automatically, criterion AIC).

Results are presented in table 6. The best model with out-of-sample RMSE criterion is ARIMA(4,1,4). On the contrary, model with included only ar4 and ma4 parameters is ranked as the best by both AIC and BIC criteria.

Table 6: Comparison of goodness-of-fit measures for ARIMA models in ETH short series.

Model	df	AIC	BIC	RMSE in sample	RMSE out of sample
ARIMA(4,1,4)	9	507.3688	526.2179	13.26062	9.195862
Included lags AR(1,4) and MA(1,2,3,4)	7	506.2194	520.8798	13.56528	9.302372
ARIMA(0,1,4) (automatically, criterion AIC)	5	505.5054	515.9771	14.82331	9.399717
Included lag 4	3	501.9709	508.2539	14.87736	9.634178

Forecasting both series using VAR model

As I have stated in the introduction, initially I have tried to create similar analysis as for ARIMA model with Vector Autoregression. However, this came out to be impossible. The reason is that the series on its full length are not cointegrated. This was proven both by running a regression BTC against ETH and testing for stationarity of residuals. To create a meaningful comparison between forecasts for VAR and ARIMA models, I had to limit the in-sample period to smaller window, that is from 2020-01-31 to 2020-03-31.

In estimating Vector Autoregressive model for bot series, I have used automatic lag order selection. All the criteria (including AIC and BIC) indicated that the best p is 1. However, after fitting that model, lag number 4 was still significant in both Bitcoin and Ethereum ACF plots.

To mitigate that effect, I have also fitted model with 4 lags. In this model, there exist no significant autocorrelations of residuals for both time series.

Table 7 below shows comparison of various goodness-of-fit statistics obtained for VAR(1) and VAR(4) models. RMSEs were obtained using one-step-ahead forecast on the same test dataset as in previous models. As can be seen, RMSE on both BTC and ETH series are lower using model with 1 lag. However, both information criteria are actually lower for model with 4 lags, despite that automatic variable selection showed the contrary. This peculiarity can be probably explained by some approximation used in VARselect function.

Table 7: Comparison of goodness-of-fit measures for VAR models

Model	df	AIC	BIC	RMSE for ETH out-of-sample	RMSE for BTC out-of-sample
p = 1	6	1312.620	1325.186	9.378566	283.0831
p = 4	18	1262.817	1299.592	10.555057	303.6446

Models comparison

Final step of my analysis was comparison of 3 approaches to modeling Bitcoin and Ethereum prices - ARIMA on 2 year period, ARIMA on shorter period, 2-month period and Vector Autoregressive model for the same shorter period.

Best models for Bitcoin prediction are:

- ARIMA model from estimation on long period - included lags 1, 9, 17, 18, 20
- ARIMA model from short period - ARIMA(1,1,0)
- VAR model from short period - 1 lag included

While for Ethereum best models are:

- ARIMA model from estimation on long period - ARIMA(20,1,20)
- ARIMA model from short period - ARIMA(4,1,4)
- VAR model for short period - 1 lag included

Visualization of one-step out-of-sample predictions for the 3 models is presented on the plot 11. For both series, all 3 types of models behave similarly - they resemble naive method forecast, that is the predictions are close to the previous value of the series.

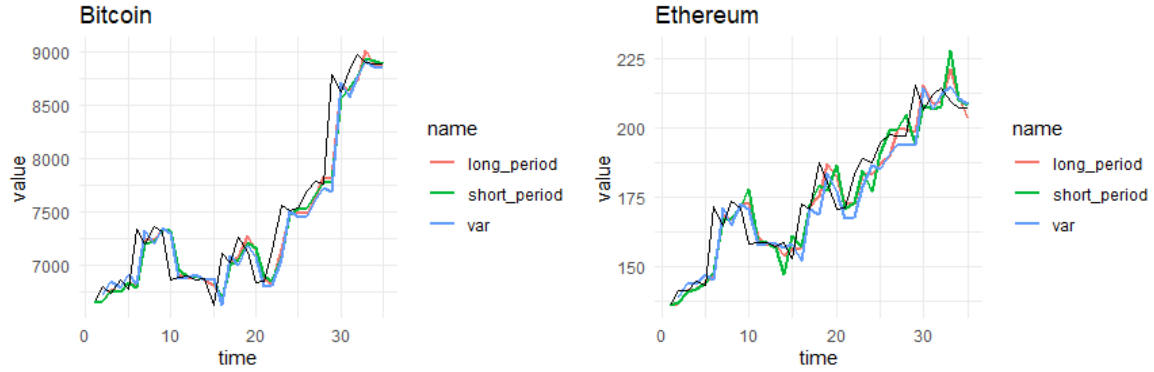


Figure 11: Forecast for BTC and ETH series obtained from ARIMA for long period and ARIMA and VAR for short period.

In the table 8, comparison of goodness-of-fit measures for Bitcoin series prediction is presented. Judging by forecast accuracy on out-of-sample period, the ARIMA model estimated with longer timespan gives the best results. Second is ARIMA model for shorter period, and the last one - VAR model. Akaike and Bayesian Information Criteria cannot be directly compared here from two reasons. One is that the first model was estimated on longer series, and second that in VAR model Information criterion is obtained for both series jointly, thus making it impossible to compare it with the value from ARIMA model.

Table 8: Comparison of the best models obtained using ARIMA and VAR methods for BTC series.

Model	df	AIC	BIC	RMSE in sample
ARIMA long period	Included lags 1, 9, 17, 18, 20	11	318.8903	257.4691
ARIMA short period	ARIMA(1,1,0) (automatically, criterion AIC)	2	495.5588	264.4412
VAR short period	p = 1	6	NA	283.0831

The same information for Ethereum series is presented in the table 9. Results of models ranking is the same as in previous case. The best performing model is the ARIMA obtained using long series, followed by ARIMA for short period and VAR model.

Table 9: Comparison of the best models obtained using ARIMA and VAR methods for ETH series.

Model	df	AIC	BIC	RMSE in sample
long period	ARIMA(20,1,20)	41	8.553134	8.319748
short period	ARIMA(4,1,4)	9	13.260622	9.195862
VAR		6	NA	9.378566

One can ask for possible explanations of such rankings. As for comparison between two ARIMA models, the one estimated on 2-year data allowed for including larger lags, which were important for the performance gain. Also, because the series was observed during longer timespan, the model became more robust to random, non-specific fluctuations that could occur in the smaller dataset.

As for the comparison between VAR and ARIMA model, there are possible sources of poorer performance of the latter. First, ARIMA model contrary to VAR includes Moving Average component, and MA lags were highly significant in both BTC and ETH models. Second, ARIMA model allows for setting some of the lags fixed at 0, contrary to VAR. This greater flexibility is probably the main reason of univariate modeling being superior to vector approach. Another one is that the second series in VAR model was probably not a very good predictor. Unfortunately, as the series are not stationary, assessing significance of coefficient in the VAR model was questionable.

Conclusions

This study was devoted to finding the best prediction of 2 cryptocurrencies prices, namely Bitcoin and Ethereum. In the analysis I have used 2 state-of-the-art forecasting models, ARIMA and VAR. Due to the fact that the series were not cointegrated over the whole time period, I have estimated ARIMA model on long period, and both ARIMA and VAR on shorter, 2-month period. Judging by quality of the forecast, I have shown that ARIMA model taking into account more observations is superior over models taking into account only the most recent past. Also, taking values of the other series into account was not helpful, as VAR model was not able to outperform ARIMA model, even when estimated on exactly the same dataset.

References

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. [OTexts.com/fpp2](https://otexts.com/fpp2). Accessed on 25.06.2020.