2019 International Conference on Data Mining Workshops (ICDMW)

Forecasting Ethereum STORJ Token Prices: Comparative Analyses of Applied Bitcoin Models

Rhonda Bush
University of Texas at Dallas
Rhonda.Bush@utdallas.edu

Soohyun Choi
University of Texas at Dallas
Soohyun.Choi@utdallas.edu

***Abstract*—The research on forecasting Ethereum STORJ token has not been widely studied compared to forecasting Bitcoin. The objective of this paper is threefold: apply existing Bitcoin price forecasting models to the Ethereum STORJ token price; evaluate the dynamics of the model predictive utility across three time horizons (h=5 days, h=20 days and h=50 days); and determine if Ethereum STORJ token clustering coefficients impact the effectiveness of the forecasting model. We choose Bitcoin forecasting models of: $ARIMA$, $ARMA$-$GARCH$,VAR, α-$Sutte$ Indicator and $NNAR$. Model effectiveness is analyzed using RMSE, MAE and MAPE.**

We find that VAR outperforms all models in the short and mid-term horizons (h=5 and h=20 days) and $NNAR$ outperforms all models in the long-term horizon (h=50 days). Non-linearity, the intrinsic value that Neural Network has, may strongly effect the forecast accuracy result. When adding the clustering coefficient to $ARIMA$, we find that the variable is significant but only marginally improves the forecasting of the Ethereum STORJ token.

The VAR model, which includes the clustering coefficient is shown to better forecast Ethereum STORJ prices in the short and mid-term forecasting horizons. The $NNAR$ model is a better model for the long-term forecasting horizon.

***Index Terms*—Cryptocurrency, Ethereum, Prediction, Forecast, STORJ**

I. INTRODUCTION

Ethereum, a decentralized application platform which includes cryptocurrency, was introduced in 2015 by Vitalik Buterin "...to create an alternative protocol for building decentralized applications, providing a different set of tradeoffs... with particular emphasis on situations where rapid development time, security for small and rarely used applications, and the ability of different applications to very efficiently interact, are important" [1]. Ethereum uses the underlying technology of blockchain, but unlike Bitcoin allows users to build applications or services (DApps), each with its own token [2]. Three of the most used DApps on the Ethereum platform are: Golem, a peer-to-peer (P2P) network for decentralized super-computing; Aurgu, a crowd-sourcing platform for prediction markets; and Bancor, a smart-token to replace currency exchanges within Ethereum [2]. Each of the DApps offers a limited number of tokens through smart contracts. DApps, allowing for P2P contracts, reduces the "middle-man" while increasing dependability as the blockchain is considered incorruptible once the smart contract is posted. The removal of third-party participants in contracts are introducing what some call Web 3.0 [3]. Smart contracts also allow for instantaneous payment upon executing the token [2].

As of August 25, 2019, Bitcoin (BTC) held 68.10 percent of the cryptocurrency market, Ethereum (ETH) held 7.53 percent, Ripple (XRP) held 4.37 percent and Storj, an Ethereum token, held 0.01 percent [4]. While Ethereum does not dominate the market, its use cases expand beyond cryptocurrencies which make it an interesting area for study. For example, the use of smart contracts has recently been adopted by the real estate industry. In 2019 the Enterprise Ethereum Alliance (EEA) published "Real Estate Use Cases for Blockchain Technology" showcasing how EEA standards, with blockchain technology "...can transform the customer experience, speed up business transactions, and develop more efficient business models across the global real estate sector" [5].

The Ethereum STORJ token is used on the Storj network, which was originally developed on the Bitcoin cryptocurrency in 2014 and migrated over to Ethereum in 2017 [6]. As of August 25, 2019, STORJ had a marketcap value of 20 million USD, with a circulating supply of 144 million out of 425 million total STORJ tokens [7]. Storj is a open source platform which provides end-to-end encrypted cloud storage services [6]. Storj allows users access computer power remotely, creating a decentralized supercomputer.

Compared to Bitcoin and other cryptocurrencies, Ethereum and Ethereum DApps analytics are substantially less developed. However, Ethereum continues to attract an increasing interest in machine learning, economics and data science communities. Among such recent results are Derbentsev et al. [8] and Glenski et al [9]. In particular, Derbentsev et al. [8] forecasted across five different time horizons for three cryptocurrencies (Bitcoin, Ehtereum and Ripple) using Autoregressive Integrated Moving Average ($ARIMA$), Autoregressive Fractionally Integrated Moving Average ($ARFIMA$) and Baysian Additive Regression Trees ($BART$). The results of Derbentsev et al. [8] show $BART$ was more accurate than $ARIMA$ or $ARFIMA$ when using RMSE. Glenski et al. [9] observed one day forecasting comparing Bitcoin, Ethereum and Monero using social signals to develop $ARIMA$ and Long Short-Term Memory ($LSTM$) neural networks. Glenski et al. [9] found that the $LSTM$ model, which incorporated social signals, slightly outperformed $ARIMA$. And found overall that the forecasting was most accurate for Bitcoin then Ethereum and Monero, using mean absolute percentage

2375-9259/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDMW.2019.00041

error (MAPE), root mean squared percentage error (RMSPE) and maximum absolute percentage error (MaxAPE) for 1 day predictions and RMSPE for predicting up to 14 days.

The objective of this paper is to develop forecasting analyses for Ethereum STORJ token data, through the following contributions:

- evaluate the utility and limitations of models, $ARIMA$ [10] [11], $ARMA-GARCH$ [12], VAR [13], α-$Sutte$ Indicator [14] [10] and Neural Network Auto-Regressive ($NNAR$) [11] model, which are conventionally used for analysis of forecasting Bitcoin prices, in application to the Ethereum STORJ token price from July 2, 2017 through March 17, 2018;
- evaluate dynamics of model predictive utility across varying forecasting horizons;
- determine if network transaction data, in the form of Ethereum STORJ clustering coefficients, impacts the effectiveness of the forecasting models.

. We use root mean square error (RMSE), mean absolute error (MAE), and MAPE to evaluate and compare the effectiveness of each model.

The remainder of this paper is organized as follows. Section 2 details related research for Bitcoin. In section 3, we describe the data. Section 4 we present the forecasting methodologies used in this study. In Section 5, we present our results and in section 6 we give our overall conclusion and next steps.

II. Related Work

While Ethereum STORJ price and Ethereum price analytics have not been widely studied, there are many developed forecasting models for Bitcoin price and volatility as well as other cryptocurrencies including those previously mentioned, Ripple and Monero. Since its implementation by Satoshi Nakamoto in January 2009, Bitcoin has the largest market capitalization compared to other cryptocurrencies, with Ethereum currently second [7].

In 2017, Bakar and Rosbi [15] found an $ARIMA(2,1,2)$ model to forecast the Bitcoin-to-USD exchange rate. Bakar and Robsi [15] confirmed that $ARIMA$ is a reliable forecasting model for Bitcoin.

In 2018, Sutiksno et al. [10] compared forecasts of Bitcoin using $ARIMA$, $NNAR$ and the α-$Sutte$ Indicator; they found that the α-$Sutte$ Indicator resulted in better accuracy over both $ARIMA$ and $NNAR$.

Munim et al. [11] employed $ARIMA$ and $NNAR$ across two samples to predict next day prices of Bitcoin. Their results indicated $ARIMA$ outperforms $NNAR$ overall, though there were samples within the testing data where $NNAR$ had a lower RMSE.

Additional methods analyzed Bitcoin through chainlets. Akcora et al. [12] used an $ARMA(2,2)$-$GARCH(1,1)$ model and an $ARMA(2,2)$-$GARCHX(1,1)$ with Bitcoin price returns and concluded that the $GARCHX$ model was observed to predict higher volatility than the $GARCH$ model and was also found to be preferable for modeling risk.

Koutmos [13] developed two separate bivariate VAR models: percent changes in Bitcoin returns and percent changes in the total number of unique Bitcoin transactions. Similar to our horizons, Koutomos also forecasted 5, 10, 20 and 30 day horizons. Overall, Koutmos found that the Bitcoin return value had a strong linkage to its transaction activity.

In 2016, McNally [16] [17] compared a recurrent neural network (RNN) and $LSTM$ network against $ARIMA$ for predicting Bitcoin price. McNally [16] [17] found that both RNN and $LSTM$ outperformed $ARIMA$. As previously mentioned, Glenski [9] found similar results with $LSTM$ outperforming $ARIMA$ for 1-day forecasting of Bitcoin, Ethereum and Monero. Jang and Lee [18] confirmed that the Bayesian Neural Network (BNN) better estimated Bitcoin prices than linear regression or support vector regression (SVR).

Greaves and Au [19] adopted several learning algorithms to predict the price of Bitcoin: Linear Regression, Logistic Regression, Support Vector Machine and Neural Networks. We previously mentioned Derbentsev et al. [8] who compared a $BART$ model to $ARIMA$ and $ARFIMA$ as well as Glenski et al. [9] who compared $ARIMA$ to $LSTM$ neural networks.

Financial and economic models tend to use historical data in development of forecasting models in addition to other economic or financial indicators. Our initial $ARIMA$ model uses historical Ethereum STORJ data, but we expand the model evaluations to include Ethereum STORJ clustering coefficient, as an additional indicator. The Ethereum STORJ clustering coefficient indicates the connectedness of nodes, or the overall network. According to Sorgente et al. [20], this could imply that if the clustering coefficient indicates a high connectedness, the network could be attractive to new users. We include the clustering coefficient as an extension of Sorgente et al. [20], that if the clustering coefficient attracts new users, it may also be an indicator of future STORJ pricing.

III. Data

Daily Ethereum STORJ token prices and the associated clustering coefficients of the transaction network are obtained for July 2, 2017 through March 17, 2018. Fig. 1 shows the historical Ethereum STORJ token prices, returns and clustering coefficients of the transaction network. The Ethereum STORJ price appears to be non-stationary whereas the clustering coefficient of the transaction network appears stationary as depicted in Fig. 1; weak stationary data is required of the time-series data for most of our models. A time-series process, $\{z_t\}_{t=-\infty}^{t=\infty}$ is weakly stationary (or covariance stationary) if $E[z_t]$ is finite and is the same for all t and if the covariance between any two observations (labeled their autocovariance), $Cov[z_t, z_{t-k}]$, is a finite function only of model parameters and their distance apart in time, k, but not of absolute location of either observation on the time scale [21].

Most non-stationary economic variables are due to trends or drifts. The random walk with drift shown in equation 1 and the stationary trend process shown in equation 2 are characterizations of most macroeconomic time series. The equations are as follows:

$$z_t = \mu + z_{t-1} + \epsilon_t, \tag{1}$$

$$z_t = \mu + \beta_t + \epsilon_t, \tag{2}$$

where z_t is the time series data value at time t, μ is drift and ϵ_t is a white noise (WN) process. Suppose we have following process with $\sigma^2 < \infty$:

$$z_t = \epsilon_t, \quad \epsilon_t \sim (0, \sigma^2) \tag{3}$$

Then, such a process with mean zero, constant variance, and no serial correlation is called zero-mean white noise or simply white noise [22]. Sometimes for short we write

$$z_t = \epsilon_t \sim WN(0, \sigma^2) \tag{4}$$

A non-stationary time-series data set is characterized by a unit root. Due to the possibility of underlying trends and drifts in the data, the true relationship between data sets is hidden [21]. Therefore, we conduct a unit root test to determine stationarity. The most widely used unit root test is Augmented Dickey-Fuller (ADF) test [23]. We evaluate the data using the ADF test and find that the Ethereum STORJ token price is non-stationary unlike the clustering coefficient of the transaction network. We use a common economic time-series data transformation, using log difference of prices (also labeled returns) to re-evaluate the stationarity of the data. The return itself has an economic insight, where the return is a complete and scale-free summary of investment opportunity and has more attractive statistical properties [24]. The log difference of prices is similar to the real return value if the return is small:

$$r_t = \frac{z_t - z_{t-1}}{z_t} \approx log(z_t) - log(z_{t-1}), \tag{5}$$

where r_t is return and z_t is the price at time t. The graph of the log difference of the Ethereum STORJ token price, titled Ethereum price return, shows a statistically significant stationary time-series plot, using the ADF test. Additionally, in financial models the historical return is the primary variable for forecasting models, as was modeled by Bakar and Munim [15] [11].

The transaction network clustering coefficients plot exhibits several dramatic spikes event though the ADF test indicates that the stationarity of the clustering coefficients are statistically significant. The spikes, in January 2018-April 2018, may be due to a January 2018 market correction in Ethereum, which caused an 18 percent drop in Ethereum's price [25]. The reduction in Ethereum's price may have impacted Ethereum STORJ transaction network. STORJ also announced in December 2017 that it was locking up 245 million STORJ tokens for 6-months as a method to control token volume and create market stability [26] [27]. Along with the announcement of a new CEO [27] during January - April 2018, any or all of these factors could have created the spikes in the transaction network. Similar to our analysis, Greaves and Au [19] claimed that the most informative feature for Bitcoin price prediction was current price. Further, Sorgente et al [20] indicated that the clustering coefficient of Bitcoin is a good indicator for overall network connectedness. Sorgente et al [20] state that if the network as a whole is highly connected, then the price is likely to increase because the transactions are distributed through the whole network and thus becoming attractive to new users. We extend both assertions assuming that the Ethereum STORJ price is the primary predictor variable and that the clustering coefficients of the transaction network could be an additional variable depending on the model. Hence, some models in this paper additionally use clustering coefficient to forecast Ethereum STORJ prices.

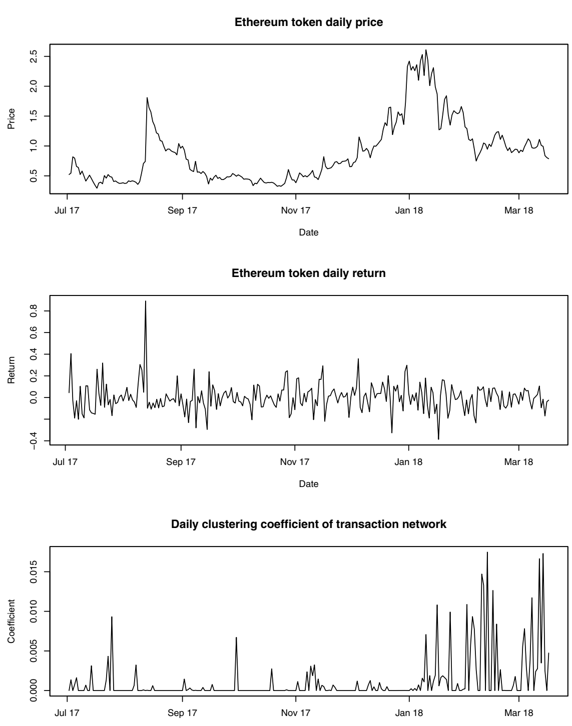


Fig. 1: Historical Ethereum STORJ token prices, returns and clustering coefficients of the transaction network

IV. METHODOLOGY

In this chapter, we investigate $ARIMA$, $ARMA$-$GARCH$, VAR, α-$Sutte$ Indicator and $NNAR$ models to evaluate Ethereum STORJ forecasting and its effectiveness. We analyze model effectiveness through RMSE, MAE and MAPE.

We use the $ARIMA$ model to benchmark Ethereum STORJ price predictions, similar to other Bitcoin studies [8] [9] [10] [11] [15]. The general $ARIMA(p, d, q)$ model is:

$$\phi_p(B)(1-B)^d Z_t = \Theta_q(B)\epsilon_t, \quad \epsilon_t \sim WN(O, \sigma^2), \tag{6}$$

where $\phi_p(B)$ is the autoregressive operator,

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p), \tag{7}$$

$\Theta_q(B)$ is the moving average operator,

$$\Theta_q(B) = (1 - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_q B^q), \quad (8)$$

and B is the backshift operator,

$$BX_t = X_{t-1}. \quad (9)$$

If $d = 0$, then $ARIMA(p, d, q)$ reduces to $ARMA(p, q)$. We follow the Box-Jenkins process to identify p, d, and q, then estimate the respective coefficients of the final model, and verify the model assumptions [28]. Likewise, $ARIMAX(p, d, q)$ model is:

$$\phi_p(B)(1 - B)^d Z_t = \alpha X_t + \Theta_q(B)\epsilon_t, \quad \epsilon_t \sim WN(O, \sigma^2), \quad (10)$$

where X_t is exogenous variable.

We then choose Akcora et al.'s [12] approach with $ARMA$-$GARCH$, where $ARMA$-$GARCH$ is an extension of $ARMA$ to include volatility measures through $GARCH$. To be specific, $ARMA(p, q)$-$GARCH(1, 1)$ is as follows:

$$z_t = \mu + \sum_{i=1}^{p} z_{t-1} + \sum_{i=1}^{q} u_{t-1} + \sigma_t u_t, \quad (11)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta\sigma_{t-1}^2, \quad (12)$$

where z_t are the observed daily returns, u_t are standard skewed Students t-innovations and σ_t is the volatility. Distribution of many financial series is skewed and has excess kurtosis [29]; the standard skewed Student's t-distribution considers such fact properly.

We also analyze the Ethereum STORJ token data using a Vector Autoregression (VAR) model, similar to Koutmos [13]. Instead of using transaction activity, we use the Ethereum STORJ clustering coefficient in addition to the price. Albeit VAR is a competent tool to conduct research of how unexpected shock effects variables simultaneously over time, the main goal in this paper is to obtain prediction values and interpret its implication. The $VAR(p)$ model is as follows:

$$z_t = c + A_1 z_{t-1} + A_2 z_{t-2} + \cdots + A_p z_{t-p} + \epsilon_t, \quad (13)$$

where A_i $(i = 1, \ldots, p)$ is a time-invariant matrix and ϵ_t is a white noise process.

We also analyze Ethereum STORJ price using Neural Network Autoregression ($NNAR$) similar to Bitcoin [16] [17]. Generally, $NNAR(p, P, k)_m$ model has inputs $(z_{t-1}, \ldots, z_{t-p}, z_{t-m}, \ldots, z_{t-mP})$ and k neurons in the hidden layer which makes neural networks non-linear [30]. Jang et al. [18] applied this methodology to Bitcoin in practice using Bayesian neural networks. We use a single layer Neural Network to develop our model.

The α-$Sutte$ Indicator was introduced by Ahmar [14] in 2015 as a new method for forecasting, originally utilized to forecast stock movement. The α-$Sutte$ Indicator is an expansion of the Sutte Indicator for use in time series data [31]. The α-$Sutte$ Indicator is as follows:

$$z_t = \frac{1}{3}\left[\alpha\frac{\Delta x}{(\alpha + \delta)/2} + \beta\frac{\Delta y}{(\beta + \alpha)/2} + \gamma\frac{\Delta z}{(\gamma + \beta)/2}\right] \quad (14)$$

where

$$\gamma = z_{t-1} \qquad \beta = z_{t-2} \qquad \alpha = z_{t-3} \qquad \delta = z_{t-4}$$

$$\begin{aligned} \Delta x &= \alpha - \delta = z_{t-3} - z_{t-4} \\ \Delta y &= \beta - \alpha = z_{t-2} - z_{t-3} \\ \Delta z &= \gamma - \beta = z_{t-1} - z_{t-2} \end{aligned}$$

and z_t is the data series observation at time t, and z_{t-k} is the data series observation at time $(t - k)$.

Forecasting is now conducted across several horizons, h: h=5 days, h=20 days, and h=50 days, which are similar to previous studies [9] [8] [13]. We choose the short term horizon to be greater than one day due to Liew et al. [32] findings that short term forecastings are "tenuous" and that "day trading cryptocurrencies may be very challenging." Thus we choose to start the horizon at under one-week and extend it almost 2-months in line with swing trading [33]. Swing trading follows the trends or "swings" of a company [34]. We classify our three horizons as short-term (h=5), mid-term (h=20), and long-term(h=20).

Various measures are consistently used to assess the predictive accuracy of forecasting models. Among those, this paper utilizes three heavily used measures based on forecast error: root mean squared error (RMSE), mean absolute error(MAE), and mean absolute percentage error (MAPE) [21] [35].

$$\begin{aligned} RMSE(h) &= \sqrt{\tfrac{1}{h}\textstyle\sum_{j=1}^{h}(e_{t+j})^2} \\ MAE(h) &= \tfrac{1}{h}\textstyle\sum_{j=1}^{h}|e_{t+j}| \\ MAPE(h) &= \tfrac{1}{h}\textstyle\sum_{j=1}^{h}|p_{t+j}| \end{aligned} \quad (15)$$

where h is the number of periods being forecasted, $e_{t+j} = z_{t+j} - \hat{z}_{t+j}$ is forecast error between forecast value $(\hat{z}_{t+j})$ and actual value (z_{t+j}), and $p_{t+j} = 100(z_{t+j} - \hat{z}_{t+j})/z_{t+j}$ is percentage error. Historically, RMSE is a popular accuracy measure, mainly due to its theoretical relevance in statistical modeling [35]. However, it is more sensitive to outliers than MAE [35]. MAPE is scale-independent unlike RMSE and MAE, but it has an extremely skewed distribution when actual value is close to zero [35].

V. Results

A. Estimation

Table I summarizes the estimation results of each model; details are explained in each subsection below.

1) ARIMA Model: Since the Ethereum STORJ price return is stationary, we develop $ARIMA(p, 1, q)$ as our baseline model. By the Akaike Information Criteria (AIC), $ARIMA(2, 1, 0)$ is shown to be the best fit[1]. AIC is chosen, among various types of information criteria, due to its considerable efficiency although it may asymptotically overestimate lag length.

When adding the clustering coefficient of the transaction network to the $ARIMA$ model as an exogenous variable,

[1]The $ARIMA$ model with drift is also considered, but its AIC value was higher than current model.

creating an $ARIMAX$ model, only marginal improvements in the predictions are shown even though the variable itself is significant. Hence, only the residual analysis and forecasting results of the $ARIMA$ model are shown in following chapters.

2) ARMA-GARCH Model: Since financial data commonly shows volatility clustering, we supplement the $ARIMA$ model with a $GARCH$ model. Thus, we apply the $ARMA$-$GARCH$ model [12], but utilize Ethereum STORJ token data to result in $ARMA(2,2)$-$GARCH(1,1)$ when applying the price return. Given the assumption of standard skewed Student's t-distribution, the estimation result is in Table I.

3) VAR Model: We next analyze the VAR model, similar to [13], for the Ethereum STORJ data. VAR variables should be stationary, thus the first variable is the Ethereum STORJ price difference and the second variable is the clustering coefficient of transaction network. The order of variables is critical in VAR modeling; the result is described in Table I. The VAR selects each of the one lagged variable by AIC as an information criteria.

4) Neural Network Model: Ethereum STORJ price estimation by the Neural Network model [30] [10] suggests one lagged input of Ethereum STORJ price and one hidden layer containing one neuron. To be specific, $NNAR(1,1)$ with average of 20 networks, each of which is a 1-1-1 network with 4 weights.

5) α-$Sutte$ Indicator: The α-$Sutte$ indicator forecasting model, provided in the sutteForecastingR package, allows for short-term horizons only. This limits our results to the h=5 case.

B. Residual Analysis

Residual analysis is crucial in testing how well a model fits the testing data and verifying that residuals are consistent with white noise. Two main factors to verify are non-serial correlation and homoscedasticity. Fig. 2 depicts the residuals from the $ARIMA$, $ARMA$-$GARCH$, VAR, and Neural Network models[2]. Residuals from $ARIMA$ and $ARMA$-$GARCH$ models appear more homoscedastic than VAR and Neural Network models.

1) ARIMA Model: The autocorrelation function (ACF), partial autocorrelation function (PACF), Bartels test [36] and Ljung-Box test [37] [38] show evidence of uncorrelated residuals.

2) ARMA-GARCH Model: The Ljung-Box test [37] [38] supports the assumption that residuals are not correlated. The Bartels test [36] cannot reject the null hypothesis of randomness.

3) VAR Model: The VAR model residual test includes two variables simultaneously: Ethereum STORJ token price difference and clustering coefficient of transaction network. The Portmanteau Test [38] alarms the validity of residual assumptions, especially of non-serial correlation.

4) Neural Network Model: The $NNAR$ model residual plots appear quite similar to those of VAR. Bartels Ratio Test [36] cannot reject the null of randomness, implying residuals are uncorrelated.

C. Forecasting

As explained earlier, the forecast horizons are classified into short-term (h=5), mid-term (h=20) and long-term (h=50). The overall forecast values are depicted in Fig. 3 with shaded area implying the forecast region. Each model's prediction values start at the shaded area, Mar 18, 2018, without update information. Fig. 4 shows accuracy measure comparison for the three different horizons. To note, the α-$Sutte$ analysis in the sutteForecastR package only provides short-term forecast values. Since forecast values using α-$Sutte$ are far higher than others, the values are excluded.

As shown in Fig 3, $ARIMA$, $ARMA$-$GARCH$, and VAR model forecast values are relatively flat with a slight downward trend compared to that of $NNAR$, which has a similar upward trend like that of the data. The α-$Sutte$ prediction mimics the previous STORJ pattern remarkably well, but is not in line with the future pattern of the Ethereum STORJ prices. Such fact implies that α-$Sutte$ is not a good forecasting model for Ethereum STORJ; $ARIMA$, $ARMA$-$GARCH$ and VAR models do not forecast as well as $NNAR$ as the horizon increases. The graphical analysis is confirmed through review of the RMSE, MAE, and MAPE measures. Since all models use the same scale, scale-dependent issues of RMSE and MAE are not problem. In the short-term and mid-term horizons RMSE, MAE, and MAPE values for VAR are the lowest with little margin. However, the power of the $NNAR$ forecast stands out in long-term with low RMSE, MAPE and MAE values. Non-linearity, the intrinsic value of $NNAR$, may strongly effect the forecast accuracy result. Our findings are consistent with McNally [16] [17] in that a neural network outperforms $ARIMA$ for Bitcoin prices. We find the opposite finding compared to Sutiksno [10], who found that α-$Sutte$ outperformed $ARIMA$ and $NNAR$ for Bitcoin prices. We find that α-$Sutte$ was least successful in forecasting Ethereum STORJ token price. Our findings are also consistent with Munim et al. [11] who stated that $ARIMA$ was a better predictor for next-day Bitcoin prices as compared to $NNAR$. Our data shows the same pattern where $ARIMA$ is a better predictor of Ethereum STORJ in the short-term horizon and $NNAR$ in the long-term.

VI. Conclusion

This paper forecasts Ethereum STORJ prices using several models: $ARIMA$, $ARMA$-$GARCH$, VAR, α-$Sutte$ Indicator and $NNAR$. The clustering coefficient of the Ethereum STORJ transaction network is a significant variable when added to the $ARIMA$ model, but including the clustering coefficient does not improve the $ARIMA$ prediction results. The VAR model employs both the Etheruem STORJ token price and the clustering coefficient. The resulting VAR model

[2]There is no residual data output given in α-$Sutte$ Indicator model developed in sutteForecastR package.

describes possible simultaneous relationships between the two factors.

In the short-term and mid-term horizons, the VAR model outperforms the others while the $NNAR$ model stands out for long term horizon forecasts; we selected these models due to their analysis of other cryptocurrency time series data.

Future model development would explore the $ARMA$-$GARCHX$ and additional Neural Network models as suggested by previous authors. This includes exploring the use of additional variables other than clustering coefficient of transaction network, for example the day of the week. Caporale and Plastun [39] found that the day of the week impacted short-term trading strategy for Bitcoin, which may impact price forecasting. Additionally, in Bitcoin analysis some features such as net Bitcoin flow, closeness centrality, mined bitcoin in the last hour, and mean node degree in the last hour are proved the informative other than price [19]. Hence, employing such similar factors in Ethereum STORJ token price is one possibility of future research. Also, if a model eliminates outliers detected in the residual plots, forecasting and estimation results are expected to improve [40].

REFERENCES

[1] J. Ray, "Welcome to the ethereum wiki!" 2019, https://github.com/ethereum/wiki/wiki.

[2] "Guides archive." [Online]. Available: https://blockgeeks.com/guides/

[3] "Whats the difference between bitcoin (btc) and ethereum (eth)?" Apr 2019. [Online]. Available: https://www.cryptocompare.com/coins/guides/what-s-the-difference-between-bitcoin-btc-and-ethereum-eth/

[4] "Storj (storj) price, marketcap, chart, and fundamentals info." [Online]. Available: https://www.coingecko.com/en/coins/storj

[5] "Real estate leaders unveil enterprise ethereum alliance blockchain use case playbook," May 2019. [Online]. Available: https://entethalliance.org/real-estate-leaders-unveil-enterprise-ethereum-alliance-blockchain-use-case-playbook/

[6] A. Batabyal, "What is storj? 2019 beginner's guide on storj cryptocurrency," Jul 2018. [Online]. Available: https://coinswitch.co/info/storj/what-is-storj

[7] "Storj (storj) price, charts, market cap, and other metrics." [Online]. Available: https://coinmarketcap.com/currencies/storj/

[8] V. Derbentsev, N. Datsenko, O. Stepanenko, and V. Bezkorovainyi, "Forecasting cryptocurrency prices time series using machine learning approach," in *SHS Web of Conferences*, vol. 65. EDP Sciences, 2019, p. 02001.

[9] M. Glenski, T. Weninger, and S. Volkova, "Improved forecasting of cryptocurrency price using social signals," *arXiv preprint arXiv:1907.00558*, 2019.

[10] D. U. Sutiksno, A. S. Ahmar, N. Kurniasih, E. Susanto, and A. Leiwakabessy, "Forecasting historical data of bitcoin using arima and α-sutte indicator," in *Journal of Physics: Conference Series*, vol. 1028, no. 1. IOP Publishing, 2018, p. 012194.

[11] Z. H. Munim, M. H. Shakil, and I. Alon, "Next-day bitcoin price forecast," *Journal of Risk and Financial Management*, vol. 12, no. 2, p. 103, 2019.

[12] C. G. Akcora, M. F. Dixon, Y. R. Gel, and M. Kantarcioglu, "Bitcoin risk modeling with blockchain graphs," *Economics Letters*, vol. 173, pp. 138–142, 2018.

[13] D. Koutmos, "Bitcoin returns and transaction activity," *Economics Letters*, vol. 167, pp. 81–85, 2018.

[14] A. S. Ahmar, "Sutte indicator: an approach to predic the direction of stock market movements." *Songklanakarin Journal of Science & Technology*, vol. 40, no. 5, 2018.

[15] N. A. Bakar and S. Rosbi, "Autoregressive integrated moving average (arima) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction," *International Journal of Advanced Engineering Research and Science*, vol. 4, no. 11, 2017.

[16] S. McNally, J. Roche, and S. Caton, "Predicting the price of bitcoin using machine learning," in *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*. IEEE, 2018, pp. 339–343.

[17] S. McNally, "Predicting the price of bitcoin using machine learning," Ph.D. dissertation, Dublin, National College of Ireland, 2016.

[18] H. Jang and J. Lee, "An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information," *Ieee Access*, vol. 6, pp. 5427–5437, 2017.

[19] A. Greaves and B. Au, "Using the bitcoin transaction graph to predict the price of bitcoin," *No Data*, 2015.

[20] M. Sorgente and C. Cibils, "The reaction of a network: Exploring the relationship between the bitcoin network structure and the bitcoin price," *No Data*, 2014.

[21] W. H. Greene, *Econometric analysis*. Pearson Education India, 2003.

[22] X. Diebold Francis, "Elements of forecasting," *Cincinnati, Thomson South Western*, 2007.

[23] C. Agiakloglou and P. Newbold, "Empirical evidence on dickey-fuller-type tests," *Journal of Time Series Analysis*, vol. 13, no. 6, pp. 471–483, 1992.

[24] R. S. Tsay, *Analysis of Financial Time Series*. John Wiley & Sons, 2005.

[25] M. Beedham, "Cryptocurrency q1 roundup: How did ethereum perform?" May 2019. [Online]. Available: https://thenextweb.com/hardfork/2019/05/01/ethereum-performance-q1-2019-etoro/

[26] B. Dale, "300 million lockup: Storj clarifies token economics in surprise reveal," Dec 2017. [Online]. Available: https://www.coindesk.com/300-million-lockup-storj-clarifies-token-economics-surprise-reveal

[27] N. De, "Former docker ceo to lead crypto-powered distributed storage startup," Mar 2018. [Online]. Available: https://www.coindesk.com/former-docker-ceo-to-lead-crypto-powered-distributed-storage-startup

[28] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[29] P. Theodossiou, "Financial data and the skewed generalized t distribution," *Management Science*, vol. 44, no. 12-part-1, pp. 1650–1661, 1998.

[30] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

[31] A. Ahmar, A. Rahman, and U. Mulbar, "α-sutte indicator: a new method for time series forecasting," in *Journal of Physics: Conference Series*, vol. 1040, no. 1. IOP Publishing, 2018, p. 012018.

[32] J. Liew, R. Z. Li, T. Budavári, and A. Sharma, "Cryptocurrency investing examined," *The Journal of The British Blockchain Association*, p. 8720, 2019.

[33] C. Mitchell, "A look at the pros and cons of day trading versus swing trading," Mar 2019. [Online]. Available: https://www.thebalance.com/day-trading-verus-swing-trading-4134456

[34] H. Pan, "A swingtum theory of intelligent finance for swing trading and momentum trading," *2004). A revised version submitted to a finance journal*, 2004.

[35] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

[36] R. Bartels, "The rank version of von neumann's ratio test for randomness," *Journal of the American Statistical Association*, vol. 77, no. 377, pp. 40–46, 1982.

[37] G. M. Ljung and G. E. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.

[38] G. E. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American statistical Association*, vol. 65, no. 332, pp. 1509–1526, 1970.

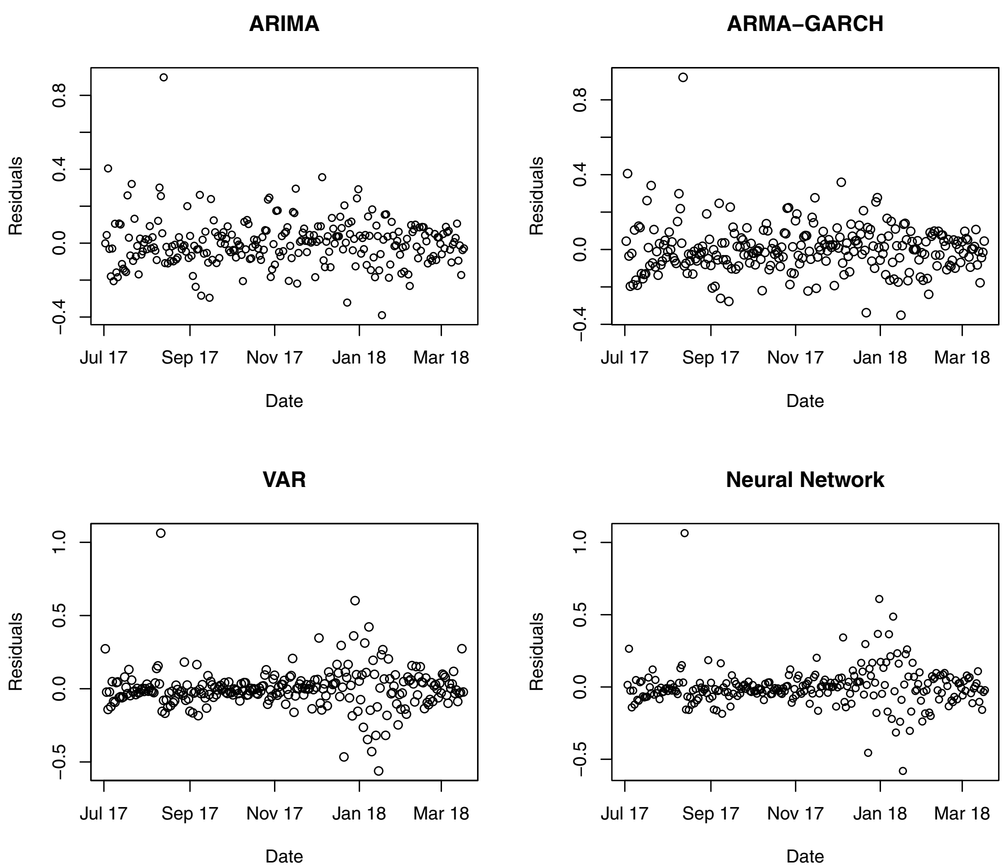
[39] G. M. Caporale and A. Plastun, "The day of the week effect in the cryptocurrency market," *Finance Research Letters*, 2018.

[40] V. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial intelligence review*, vol. 22, no. 2, pp. 85–126, 2004.

TABLE I: Estimation equation: ARMA, ARMA-GARCH, VAR, Neural Network, α-Sutte

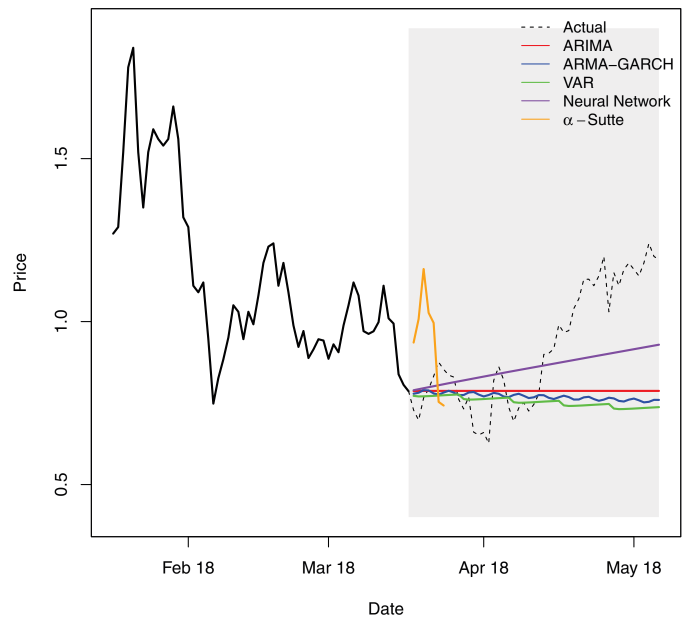
ARIMA	$y_t = 0.0116\ y_{t-1} - 0.0254\ y_{t-2} + u_t,$ or $y_t = 0.0115\ y_{t-1} - 0.0287\ y_{t-2} + 1.6778\ x_t + u_t$
ARMA-GARCH	$y_t = -0.0008 + 0.4087\ y_{t-1} - 0.9901\ y_{t-2} - 0.3907\ u_{t-1} + 1.0004\ u_{t-2} + \sigma_t u_t,$ $\sigma_t^2 = 0.0000 + 0.0000 u_{t-1}^2 + 0.9985\ \sigma_{t-1}^2$
VAR	$\begin{pmatrix} \Delta Y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0.0058 \\ 0.0009 \end{pmatrix} + \begin{pmatrix} -0.0200 & -4.2975 \\ -0.0023 & 0.1767 \end{pmatrix} \begin{pmatrix} \Delta Y_{t-1} \\ x_{t-1} \end{pmatrix} + u_t$
Neural Network	No equation
α-Sutte	No equation

[a] y_t is Ethereum STORJ price return, Y_t is Ethereum STORJ price, x_t is clustering coefficient of transaction network, and u_t is residuals.



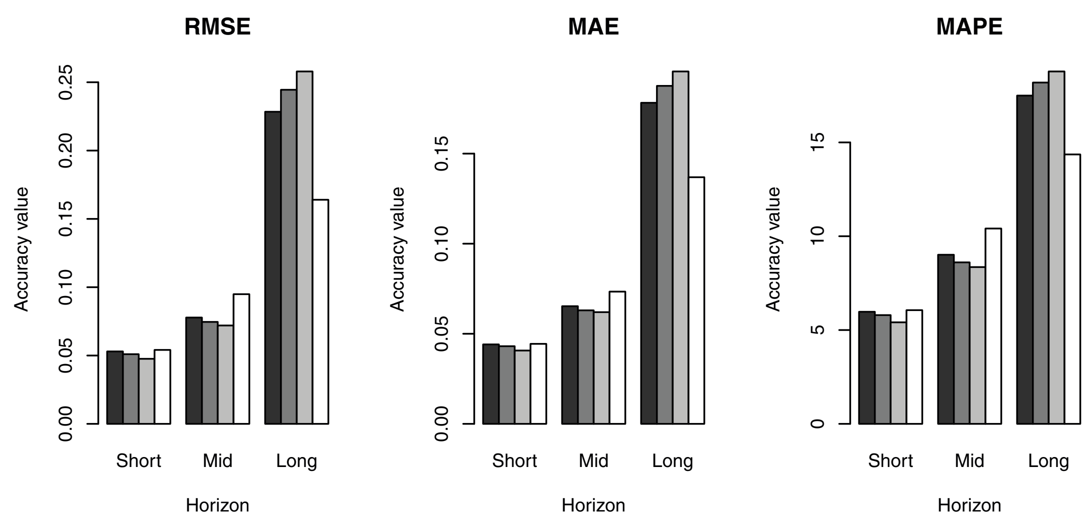
Note: The residuals are derived based on estimation period: July 2, 2017 through March 17, 2018.

Fig. 2: Residuals: ARIMA, ARMA-GARCH, VAR, Neural Network



Note: In the shaded area, the dotted line indicates actual Ethereum STORJ prices, and colored lines represent forecasting values derived by the ARIMA model (red), ARMA-GARCH model (blue), VAR model (green), Neural Network model (purple), and α-$Sutte$ model (orange). The forecasting period ranges from Mar 18, 2018 to May 6, 2018.

Fig. 3: Ethereum STORJ token price forecast



Note: On horizontal axis, "Short" indicates 5-day horizon, "Mid" indicates 20-day horizon and "Long" indicates 50-day horizon. The forecasting period ranges from Mar 18, 2018 to May 6, 2018. For the definition of RMSE, MAE, and MAPE, refer to Equation (15)

Fig. 4: Accuracy measure comparison by different horizon