# Time Series Analysis and prediction of bitcoin using Long Short Term Memory Neural Network

Temiloluwa I. Adegboruwa [1*] Steve A. Adeshina [2*] & Moussa M. Boukar [3]
[1,2&3]Department of Computer Science, Nile University of Nigeria
email : steve.adeshina@nileuniiversity.edu.ng

**Abstract**—Bitcoin is the first digital currency that uses decentralization to solve the issue of trust in performing the functions of a digital currency successfully. This digital currency has shown extraordinary growth and intermittent plunge in value and market capitalization over time. This makes it important to understand what determines the volatility of bitcoin and to what extent they are predictable. Long Short Term Memory Neural Networks (LSTM-NN) have recently grown popular for time series prediction systems but there has been no consensus on methods to model time series inputs for LSTMs, this paper proposes the need for this problem to be solved by conducting an experimental research on the efficacy of an LSTM-NN given the form of its time-series input features.

**Index Terms**—Long Short Term Memory Neural Networks, Bitcoin, Time Series, detrend.

## 1 Introduction

Bitcoin is a digital currency that has emerged has a powerful safe haven and has been known to be a hedge over other financial assets [1]. Before bitcoin, there had been many attempts to create a digital currency but centralization was an issue, so new projects tried decentralization by implementing it on a permission-less peer to peer (P2P) network, where anyone can be part of the network. But that came with its own problem of trust and security. So bitcoin proposed an ingenious protocol to solve this known as the Proof of Work (POW) where write access is only given to nodes that have spent the most power in the network [2]. Furthermore, Bitcoin uses cryptography to preserve integrity and authenticity. Bitcoin has gone on to fulfil the functions of a currency and increased in value over the years more than any other financial asset. However, This asset turns out to be a very volatile one which increases its risk tremendously. Therefore, attempting to predict it can be challenging and rewarding. This paper is a time series analysis on bitcoin and its related variables, this paper also addresses a gap in neural networks prediction of a time series problem. There has been no known consensus on methods to model time series variables for neural networks prediction, from our preliminary literature review, it is either not discussed or the variables are given to the neural network in their raw form. Time series variables are known to generally possess a unit-root problem (A problem where the mean and standard deviation of the series deviates from 0 and never reverts) to solve this for conventional time series forecasting models like Autoregressive Integrated Moving Average (ARIMA), Autoregressive (AR), Generalized Autoregressive conditional heteroscedasticity (GARCH), the problem has to be removed by differencing or detrending. In this paper we address this conundrum by conducting an experiment on the performance of an LSTM-NN given the form of its input.

### 1.1 Background and Motivation

### 1.2 Literature Review

Research conducted on the variables in the Bitcoin Network has shown that it's volatility is predictable Xu et al [3]. So there is a need to understand what variables serves as probable catalyst to the increase in volatility of bitcoin. Some of the variables known to either move with bitcoin, move before bitcoin or move after bitcoin have been speculated by researchers over the years. Giaglis et al [4] found a positive correlation with mining hash rate and bitcoin supply and a negative correlation with S&P 500 and EUR/USD. This shows the relationship between bitcoin and the economy of some countries. E. Bouri et al [5] showed that the Global Financial Stress Index (GFSI) have a left and right tail effect on bitcoin returns (BTC) distribution. E. Demir et al [6] also showed the impact of increase in attention using the same method used by E. Bouri et al. Increase in transaction cost and unconfirmed transactions in the network has also been shown to have notable effect on bitcoin returns D. Easley et al [7]. Neural Network models have been proposed to predict the returns of Bitcoin in recent researches. Support Vector Machines (SVMs) have been shown to capture transaction cost effect and Artificial Neural Networks (ANN) were shown to take advantage of short run informational inefficiencies which shows arbitrage potential H. Kimura et al [8]. LSTMs were shown to perform better than Generalized Regression Neural Network (GRNN) S. Lahmiri and S. Bekiros [9].

### 1.3 Related works, Literature Gap and proposed method

The conventional way to create a forecasting model for time series in econometrics is to first test the time series for unit roots and remove the unit root by differencing or detrending

978-1-7281-5160-1/19/31.00 © 2019 IEEE

the time series. The question is should this also be considered for Neural Network and Deep Learning models. There seem to be a lack of consensus on modeling time series variables before training them on an NN model. W. C. Hung et al [10] proposed a model where autoregression is added as a feature which performed better than an LSTM without autoregression. M. Butler and D. Kazakov [11] found that non-stationarity can affect learning process of an ANN and can get worse depending on the forecast horizon. R. H. Brown et al [12] trained a forecasting model with using their own proposed detrending method to detrend variables and achieved better forecasting. So our paper addresses the gap of the choice of modeling the input features of an LSTM.

## 2 Research Design

### 2.1 Research Process

The research design involves various steps. The first step is data collection. Data for the research was collected through the Quandl API, making it live. The date was from 16-09-2010 to 21-05-2019. A total of 3170 samples for 11 variables. The second is to conduct a stationarity test on all the series to determine their status, the third is to then difference and detrend series found to possess unit root, the fourth is to then carry out some tests to determine relationships, the fifth step is to use the findings from all the previous steps to model two LSTMs and make inferences from the MSE, MAE, and R2 result of the two models. Figure 1 shows the research process.

## 3 STATIONARITY AND UNIT-ROOT TESTS

### 3.1 Stationarity

A stationary series is one without the existence of trend or seasonality, while a non-stationary series is one that has any of the four components of a time series: Trend, Seasonality, Cyclical pattern, Irregularity [13]. Thus, a time series is stationary if its mean, variance and auto-covariance are independent of time. Consider a time series $Y_t$ and a first order autoregressive model AR(1) with an intercept term given in Eq 1:

$$\begin{gathered} y_t = a_0 + a_1 y_{t-1} + e_t \\ if\ a_1 = 1\ then\ eqn\ 1.\ simplifies\ to: \\ y_t = a_0 + y_{(t-1)} + e_t \end{gathered} \tag{1}$$

where $y_t$ is the observation of the series at time t, $a_0$ is a drift term, $a_1$ is the constant term and $e_t$ is the residual. This is a non-stationary process with intercept, and any shock to the system will have a permanent effect. But if $a_{1<1}$ then any shock on the time series will eventually return to its stable trend making it a stationary process. To illustrate further, The expectation of a non-stationary time series at time period t is:

$$E[X_t] = E\left[\sum_{t=1}^{t} Z_t\right] = \sum_{t=1}^{t} E[Z_t] = \mu t \tag{2}$$

Where $E[X_t]$ is the expected value of the time series at point $t, Z_t$ is the random noise associated with point i, $\mu$ . This shows that the expectation changes by location and it possesses a systematic change in mean

### 3.2 Augmented-Dickey Fuller Test

In Time Series Analysis, it is important to check for unit root problems in the variable so that the statistical models that relies on constant mean and variance would be reliable. To test the series for stationarity a method proposed by D. Dickey and W. Fuller [14] known as the Augmented Dickey Fuller (ADF) test is used. The ADF tests the null hypothesis that the process is non-stationary. The ADF returns a value between -1 and 1. The null hypothesis is accepted if the ADF is close to 1 and rejected if far from 1. A test statistic is used to compute a confidence level using critical levels.

## 4 Proposed Method

### 4.1 Detrend: Removing trend from a series

Our proposal is to detrend a series that has been found to possess a unit root and use the output of the detrend function to train an LSTM. This method is adopted to engineer time series features. To detrend a series, we use Linear regression to compute the regression residuals of a series [15]. The algorithm to remove a trend from a series is given in Algorithm 1. The variable $\hat{D}$ is the new values of the series after it has been detrended. To illustrate this process, consider a time series Y:

$$\begin{aligned} Y_o = 0 \to Y_1 = Y_0 + e_1 \to Y_2 = e_1 + e_2 \cdots \to Y_t \\ = e_1 + e_2 + \cdots + e_t \end{aligned} \tag{3}$$

Where $e_t$ is the residual at time t. A linear regression for this process uses $Y_{(t-1)}$ as the explanatory variable and computes the error of the prediction of $Y_t$. The error therefore becomes the new series detrended.

## 5 Relationships between Lags and explanatory variables

We proceed to compute correlation coefficient for all the variables in the research by calculating Pearson r correlation coefficient for continuous variables and Spearmans rank correlation coefficient for discrete variables. We also compute Partial autocorrelation function for the target variable i.e. Bitcoin returns (BTC) and its relationship with its lagged values

## 6 LSTM-NN

The Long Short Term Memory Neural Network is a variant of Recurrent Neural Networks proposed to solve the issue of vanishing gradient [16]. The design has a basic unit called the memory cell, which has its own fixed- weight self-connection and a candidate for the new value of the cell. The design also consist of three multiplicative gate units, the update, output and forget gate. This gates learn to allow and deny access to constant error flows. Thus, mitigating the vanishing and exploding gradient problems. Figure 2 is a diagram that depicts how the LSTM computes a series. The computation details for the LSTM is given below:

$$candidateCELL = \hat{C_t^k} = ELU(W_c[a_t^{k-1}, x_t^k] + b_c) \tag{4}$$

$$UpdateGate = U_t^k = ELU(W_u[a_t^{k-1}, x_t^k] + b_u) \quad (5)$$

$$Outputgate = f_t^k = ELU(W_f[a_t^{k-1}, x_t^k] + b_f) \quad (6)$$

$$Outputgate = O_t^k = ELU(W_0[a_t^{k-1}, x_t^k] + b_0) \quad (7)$$

$$memorycell = C_1^k = O_1^k.\hat{C_t^k} + f_t^k.C_t^{k-1} \quad (8)$$

$$Newactivationunit = a_t^k = O_t^k \ . \ ELU(C_t^k) \quad (9)$$

$$y_t^k = linear(a_t^k) \quad (10)$$

where $\hat{C_t^k}$ the candidate input for the memory cell, $U_t^k$ , $f_t^k$ and $O_t^k$ are the gates to determine boolean variables that controls the activation of the memory cell $C_1^k$ of each unit of the model. $a_t^k$ is the activation output of each unit and $y_t^k$ is the predicted output of each unit

**Algorithm 1: Algorithm to detrend series**

**Require:** m ← coefficient term
**Require:** b ← intercept term
Series ← *Time series data*
Ď ← *emptylist*
Ý ← *empty list*
*X* ← *lag*1 *of series*
Y ← *value in series*
**for** i in X
  $\hat{Y}_t = m\, X_i + b$
  $\check{D} = Y_t - \hat{Y}_t$
**End**
**return** Ď

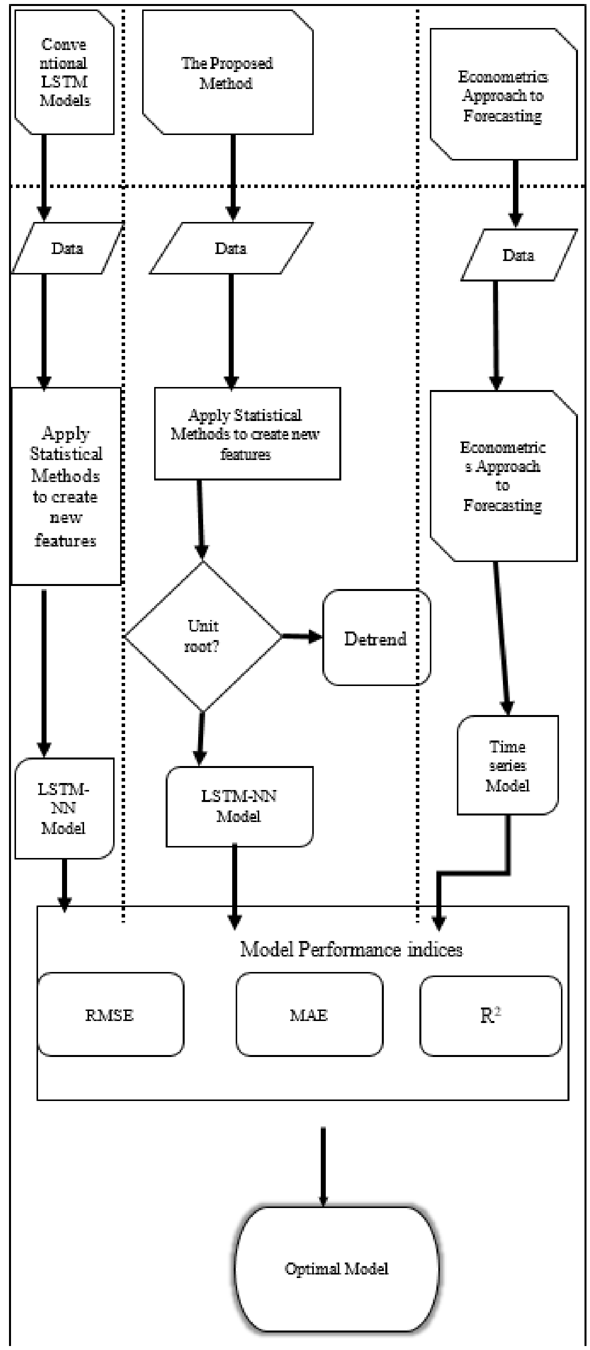


Fig. 1: Research Design for new method compared to other methods

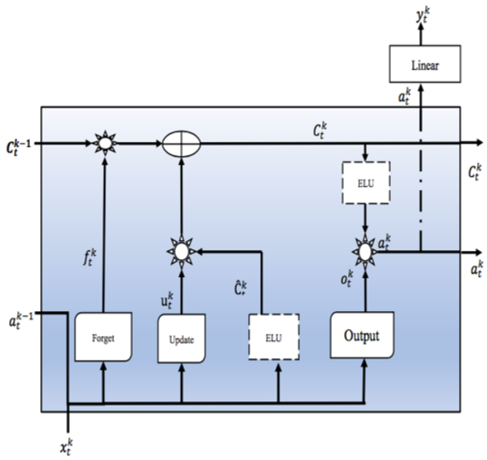


Fig. 2: Representation of the LSTM Kernel

proposed method also works better than the most recently proposed method for forecasting bitcoin by W. C. Hung *et al.*

TABLE 1: Correlation coefficient tab between all the variables

| | BTC | TC | MR | VOL | HR | SP |
|---|---|---|---|---|---|---|
| **BTC** | 1 | 0.82 | 0.96 | 0.83 | 0.93 | 0.93 |
| **TC** | 0.82 | 1 | 0.83 | 0.61 | -0.22 | -0.24 |
| **MR** | 0.96 | 0.83 | 1 | 0.88 | -0.27 | -0.28 |
| **VOL** | 0.83 | 0.61 | 0.88 | 1 | -0.36 | -0.37 |
| **HR** | 0.93 | 0.36 | -0.27 | 0.1 | 1 | 0.99 |
| **SP** | 0.93 | -0.24 | -0.28 | -0.37 | 0.99 | 1 |

## 7 Hyper-parameter Tuning, Hard & Software Environment and Performance Evaluation Metrics

### 7.1 Hyperparameters choice

In this paper, we use the Mean Squared Error as the loss function for training the model, we use the Exponential Linear Unit (ELU) activation function [17] to activate units in the LSTM. The Adamax optimizer [18] was used to optimize the loss function and find the most appropriate weights for the units in the model. The LSTM had 12 units, 1000 epochs, 100 batches

### 7.2 Software and Hardware Environment

The keras frame work was used to design the model and the model was run on the spyder python environment. The research was done on a MacBookpro 2011, core i7, 8G RAM.

### 7.3 Performance Evaluation Metrics

The models were evaluated using The Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and R Squared ($R_2$)

## 8 Result and Discussion

The result of the correlation analysis is given in table 1, it reveals that bitcoin returns (BTC) comoves with supply, Network hash rate, Transaction volume, and transaction cost. The full result of the analysis is given in table 1. The result also substantiates our observation of the halving periods effect on the volatility of bitcoin. The LSTM model result is given in table 2. The data was split into training and testing sets to evaluate prediction performance. The proposed model trained had a better train and test error, also had a better RMSE, MAE, R squared. The line plot for the prediction and actual observation is given in figure 3 and 4. And the train test error diagnosis is given in figure 5 and 6 The results of this research shows that Neural Network time series prediction models perform better when trend-observed explanatory variables are detrended. The

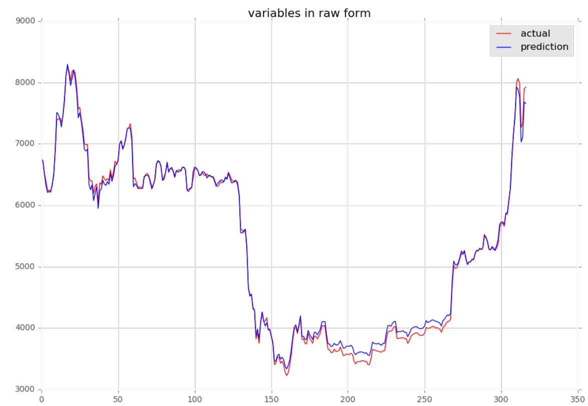


Fig. 3: actual observation and prediction line plot for model trained with variables in raw form

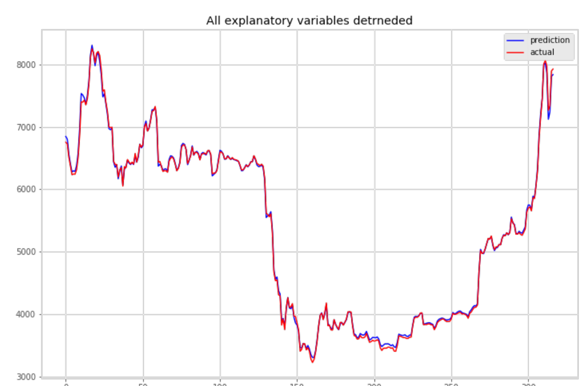


Fig. 4: all explanatory variable

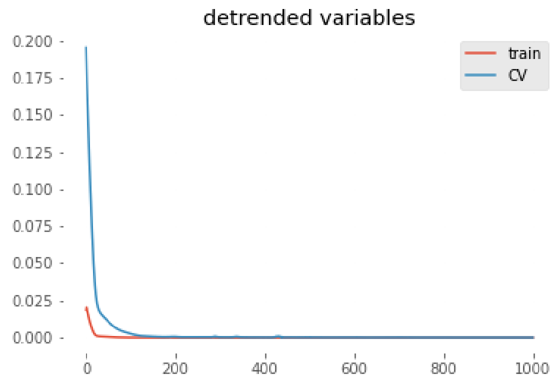


Fig. 5: Error diagnostic for Model trained with detrended values

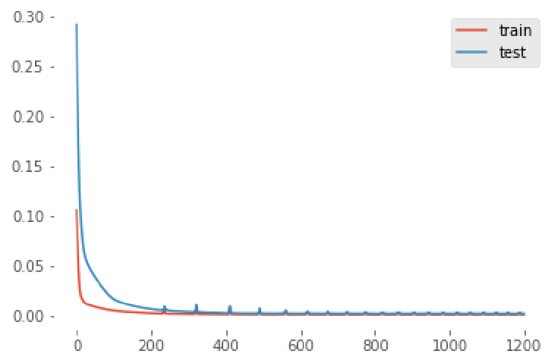


Fig. 6: Error diagnostic for Model trained with raw values (Benchmark)

| | LSTM 1 Model (Variables in raw form) | LSTM 2 Model (Detrended variables) | **W. C. Hung et al** |
|---|---|---|---|
| RMSE | 0.016 | 0.006 | 0.022 |
| MAE | 0.011 | 0.003 | 0.017 |

Fig. 7: Result of the experiment

## 9 Discussion & Conclusion

From the result obtained in this research, We can conclude that just as important as it is for an analyst to conduct several statistical tests and conclude on whether to transform series or not in a research, it is also important to conduct these tests and consequent transformations for a neural network to perform better in a time series problem. To extend and improve this research, it would be great to build a sequence to sequence LSTM, where all the variables are forecasted for a given time horizon and an Autoregressive Distributed Lag model can be used to capture the auto-correlation of all the variables in this research. Other detrending methods can also be tried, perhaps the one proposed by R. H. Brown et al can be used.

### 9.1 Recommendation

In summary, this paper proposes a new method of modeling time series for neural networks. The results obtained from this research has shown that the method works and can be applied in time series problem for neural networks.

### 9.2 Acknowledgment

We would like to thank Nile University of Nigeria for their support during this research, we will also like to thank our families and friends that made this research possible. Special thanks goes to Mr Yinka and Mr Ahmed Adeniyi for their maximum support during this research.

## References

[1] *E. Bouri, S. Jawad, D. Roubaud, "Cryptocurrency as hedges and safe-havens for US equity sectors," The Quarterly Review of Economics and Finance., 2019.*

[2] *N. Satoshi, "Bitcoin: A peer-to-peer electronic cash system"., 8 January 2008. [Online]. Available: http://bitcoin.org/bitcoin.pdf.. [Accessed 2016 June 2].*

[3] *G. Xu and Nhi N. Y. Vo., "The volatility of Bitcoin returns and its correlation to financial markets,," in 2017 International Conference on Behavioral, Economic, Socio-cultural Computing, Krakow, 2017.*

[4] *G. M. Giaglis, D. N. Sotiropoulos, C. Bilanakos, D. Pournarakis and I. Georgoula, "Using Time Series Analysis and Sentiment Analysis to detect the determinant of bitcoin prices," Mediterranean Conference on Information systems, SSRN, vol. 69, pp. 297-307, 2015.*

[5] *E. Bouri, R. Gupta, C. Kueng, M. Lau, D. Rouband and S. Wang, "Bitcoin and global financial stress: A copula-based approach to dependence and causality in the quantiles," The Quarterly Review of Economics and Finance, vol. 69, no. 10, pp. 297-307, 2018.*

[6] *E. Demir, S. Dastgir, G. Downing and G. Gozgor, " The causal relationship between Bitcoin attention and Bitcoin returns: Evidence from the Copula-based Granger Causality Test," Financial Research Letters, no. 28, pp. 160-164, 2019., vol. 12, no. 28, pp. 160-164, 2019.*

[7] *D. Easley, M. O'Hara and S. Basu, "From mining to markets: The evolution of bitcoin transaction fees," Journal of Financial Economics, 2019.*

[8] *H. Kimura, M. J. Silva de Souza, F. W. Almudhaf, B. M. Henrique, S. A. B. Negredo, R. D. G. Franco and V. A Sobreiro, "Can Artificial Intelligence Enhance the Bitcoin Bonanza," The Journal of Finance and Data Science, vol. 5, no. 2, pp. 83-98, 2019.*

[9] *S. Lahmiri and S. Bekiros, " Cryptocurrency forecasting with deep learning chaotic neural networks," Chaos, Solitions and Fractals, vol. 118, pp. 35-40, 2019.*

[10] *W. Chih-Hung, C. C. Lu, M. Yu-Feng and L. Ruei-Shan, "A New Forecasting Framework for Bitcoin Price with LSTM.," in 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore, 2018.*

[11] *M. Butler and D. Kasakov, "The effects of variable stationarity in a financial time-series on artificial neural networks," in 2011 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFER), Shenzhen, 2011.*

[12] *R. H. Brown, S. R. Vitulo, C. F. Corliss, M. Adya, P. E. Kaefer, and R. J Povinelli. , "Detrending daily natural gas consumption series to improve short-term forecasts.," in 2015 IEEE Power & Energy Society General Meeting, Denver* 2015

[13] *R. S. Tsay, "Stationarity," in Analysis of Financial Time Series second edition, Chicago, John Wiley and Sons Inc., 2005, pp. 72-75.*

[14] *Dickey, David, and Wayne A. Fuller, "Distribution of the estimators for time series regressions with a unit root," Journal of the American Statistical Association, vol. 74, no. 366, pp. 427-431, 1979.*

[15] *A Carlson and W. R. Jennifer, "Regression Analysis," in Introduction to Statistics: An active Learning Approach, Los Angeles, SAGE, 2018, pp. 1459-1473.*

[16] *H. Sepp, and J. Schmidhuber, "LSTM can solve hard long time lag problems," Advances in Neural Information Processing Systems, vol. 10, no. 20, pp. 473-479, 1997*

[17] *C. D. Arne, T Unterthiner, and S Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," Johannes Kepler University, Linz, 2015*

[18] *D. P. Kingma and B. J. Lei, "Adam: A method for stochastic optimization," University of Toronto, Toronto, 2014.*