AN EMPIRICAL STUDY ON MODELLING AND PREDICTION OF BITCOIN PRICES WITH BAYESIAN NEURAL NETWORKS BASED ON BLOCKCHAIN INFORMATION

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

ROSE MARY ABRAHAM

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

AN EMPIRICAL STUDY ON MODELLING AND PREDICTION OF BITCOIN PRICES WITH BAYESIAN NEURAL NETWORKS BASED ON BLOCKCHAIN INFORMATION

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

ROSE MARY ABRAHAM

Department of Computer Science & Engineering

Mar Athanasius College Of Engineering Kothamangalam

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MAR ATHANASIUS COLLEGE OF ENGINEERING
KOTHAMANGALAM

CERTIFICATE

This is to certify that the report entitled **An Empirical Study On Modelling And Prediction Of Bitcoin Prices With Bayesian Neural Networks Based On Blockchain Information** *submitted by* **Ms. ROSE MARY ABRAHAM**, *Reg. No.* **MAC15CS051**, *towards partial fulfillment of the requirement for the award of Degree of Bachelor of Technology in Computer science and Engineering from APJ Abdul Kalam Technological University for December 2018 is a bonafide record of the seminar carried out by her under our supervision and guidance.*

.....	.....	.....
Prof. Joby George	**Prof. Neethu Subash**	**Dr. Surekha Mariam Varghese**
Faculty Guide	***Faculty Guide***	***Head of the Department***

Date: Dept. Seal

ACKNOWLEDGEMENT

First and foremost, I sincerely thank the God Almighty for his grace for the successful and timely completion of the seminar.

I express my sincere gratitude and thanks to Dr. Solly George, Principal and Dr. Surekha Mariam Varghese, Head Of the Department for providing the necessary facilities and their encouragement and support.

I owe special thanks to the staff-in-charge Prof. Joby George , Prof. Neethu Subash and Prof. Joby Anu Mathew for their corrections, suggestions and sincere efforts to co-ordinate the seminar under a tight schedule.

I express my sincere thanks to staff members in the Department of Computer Science and Engineering who have taken sincere efforts in helping me to conduct this seminar.

Finally, I would like to acknowledge the heartfelt efforts, comments, criticisms, co-operation and tremendous support given to me by my dear friends during the preparation of the seminar and also during the presentation without whose support this work would have been all the more difficult to accomplish.

ABSTRACT

Bitcoin, the worlds first decentralized cryptocurrency, has recently attracted attention in the fields of economics, cryptography, and computer science. There have been previous studies for the prediction of bitcoin prices using linear modelling and machine learning techniques, but a machine trained with these could exhibit only poor performance. This work uses Bayesian Neural Networks to predict the bitcoin prices based on blockchain information. The study first describes a linear regression model for prediction. The results of this model are then compared with the results obtained from using the BNN model. The BNN model takes in blockchain information and other variables as input and gives bitcoin log price and volatility as response variables. The performance was measured using Root Mean Square Error(RMSE) and Mean Absolute Percentage Error(MAPE)methods. The BNN model outperformed the linear-regression-model and support-vector-regression(SVR)models. It also succeeded in relatively accurate direction prediction in terms of bitcoin price volatility.

Contents

List of Figures

List of Abbreviations

BNN	Bayesian Neural Network
ANN	Artificial Neural Network
RNN	Recurrent Neural Network
MLP	MultiLayer Perceptron
SVR	Support Vector Regression
VIF	Variation Inflation Factor
PoW	Proof of Work
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
LSTM	Long Short Term Memory
ARIMA	Autoregressive integrated moving average
GARCH	Generalized Autoregressive Conditional Heteroskedasticity

Introduction

Bitcoin is a successful cipher currency introduced into the financial market based on its unique protocol and Nakamotos systematic structural specification. Unlike existing fiat currencies with central banks, Bitcoin aims to achieve complete decentralization. Participants in the Bitcoin market build trust relationships through the formation of Blockchain based on cryptography techniques using hash functions. Inherent characteristics of Bitcoin derived from Blockchain technologies have led to diverse research interests not only in the field of economics but also in cryptography and machine learning.

The study, performs a practical analysis on modeling and predicting of the Bitcoin process by employing a Bayesian neural network (BNN), which can naturally deal with increasing number of relevant features in the evaluation is conducted. A BNN includes a regularization term into the objective function to prevent the overfitting problem that can be crucial to our framework.

When the machine considers a lot of input variables, a trained machine can be complex and suffer from the overfitting problem. BNN models showed their effect to the financial derivative securities analysis.Formation of Blockchain, a core technology of Bitcoin, distinguishes Bitcoin from other fiat currencies and is directly related to Bitcoins supply and demand. To the best of our knowledge, in addition to macroeconomic variables, direct use of Blockchain information, such as hash rate, difficulties, and block generation rate, has not been investigated to describe the process of Bitcoin price.

To fill this gap, the current study systematically evaluates and characterizes the process of Bitcoin price by modeling and predicting Bitcoin prices using Blockchain information and macroeconomic factors. We also try to account for the remarkable recent fluctuation, which is shown in Figure 1 and has not been considered in previous studies.

Related Works

Numerous studies have been conducted recently on modeling the time series of Bitcoin prices as a new market variable with specific technical rules. Relatively few studies have thus far been conducted on estimation or prediction of Bitcoin prices.

- Generalized Autoregressive Conditional Heteroskedasticity (GARCH) volatility analysis is performed to explore the time series of Bitcoin price [1].
- Various studies on statistical or economical properties and characterizations of Bitcoin prices refer to its capabilities as a financial asset; these research focus on statistical properties, inefficiency of Bitcoin according to efficient market hypothesis, hedging capability speculative bubbles in Bitcoin , the relationship between Bitcoin and search information, such as Google Trends and Wikipedia , and wavelet analysis of Bitcoin.
- Relatively few studies have thus far been conducted on estimation or prediction of Bitcoin prices. A reference study evaluates Bitcoin price formation based on a linear model by considering related information that is categorized into several factors of market forces, attractiveness for investors, and global macro-financial factors [2].
- Another work predicts the Bitcoin pricing process using machine learning techniques, such as recurrent neural networks (RNNs) and long short-term memory (LSTM), and compare results with those obtained using autoregressive integrated moving average (ARIMA) models [3].

There are few practical and systematic empirical studies on the analysis of the time series of Bitcoin.

Proposed Framework

3.1 Overview of bitcoin

This section introduces the concept of Bitcoin and the economics of Bitcoin. It also elaborates the factors on which Bitcoin price depends.

3.1.1 Bitcoin

Bitcoin is a cryptocurrency, a form of electronic cash. It is a decentralized digital currency without a central bank or single administrator that can be sent from user-to-user on the peer-to-peer bitcoin network without the need for intermediaries.

Transactions are verified by network nodes through cryptography and recorded in a public distributed ledger called a blockchain. Bitcoin was invented by an unknown person or group of people using the name Satoshi Nakamoto and released as open-source software in 2009. Bitcoins are created as a reward for a process known as mining. They can be exchanged for other currencies, products, and services.

Bitcoin provides the following advantages over normal fiat currency:

- **Decentralized power:** Every monetary system is controlled by a central governing authority, like banks. With Bitcoins, this system got distributed and decentralized with everyone who is a part of Bitcoin system, and it keeps on growing.

- **Public ledger:** Most banking and financial organizations follow a private ledger. Only thing a customer is aware of when he puts his money into this is whether or not there are certain transactions that he have made as a person as such. He has no idea whether someone else has taken this money or the bank has invested this money somewhere else and so forth. However, with Bitcoin, the ledger itself is public. Everyone who becomes

part of this blockchain network gets a copy of the entire blockchain as soon as he sign up.Although blockchain provides the complete details with respect to transactions that happens as part of network, the person cant still know who is doing the transactions and this is the anonymity that blockchain gives.

- **Immutable to hacks:** Blockchain system, the backbone of Bitcoin system, is completely immutable to hacks. Any transaction that takes place cannot be modified ahead and even if you try to modify it, the blockchain system is built so securely that the transaction detail gets rejected.

- **No double spending:** Most financial organizations that have a digital platform have faced the problem of double spending. Through blockchain system, double spending is not possible and this is because of how the blockchain system is structured and created.

3.1.2 Economics of bitcoin

Bitcoin is a digital asset designed to work in peer-to-peer transactions as a currency. Bitcoins have three qualities useful in a currency, according to The Economist in January 2015: they are ”hard to earn, limited in supply and easy to verify”.However, as of 2015 bitcoin functions more as a payment processor than as a currency.

An important determinant of Bitcoin price is the interaction between Bitcoin's supply and demand. The supply of Bitcoin determines the amount of units in circulation and thus its scarcity on the market.The total Bitcoin supply, S_B, is represented by

$$S_B = P_B B \qquad \text{(Equ:3.1)}$$

where P_B denotes the exchange rate between Bitcoin and dollar (i.e. dollar per unit of Bitcoin), and B is the total capacity of Bitcoins in circulation.

The demand of Bitcoin is mainly determined by transaction demand as a medium of exchange , by value in future exchange. The total Bitcoin demand depends on the general price level of goods or services, P; the economy size of Bitcoin, E; and the velocity of Bitcoin, V,

which is the frequency at which a unit of Bitcoin is used for purchasing goods or services. The total demand of Bitcoin, D_B, is described as followed by:

$$D_B = \frac{PE}{V} \tag{Equ:3.2}$$

The market equilibrium with the perfect market assumption is acquired when the supply and the demand of Bitcoin is the same amount. The equilibrium is therefore achieved at

$$P_B = \frac{PE}{VB} \tag{Equ:3.3}$$

This equilibrium equation implies that in the perfect market, the Bitcoin price in dollars is affected proportionally by the general price level of goods or services multiplied by the economy size of Bitcoin, and inversely by the velocity of Bitcoin multiplied by the capacity of the Bitcoin market.

Fig.3.1 shows the time series of Bitcoin price obtained from https://bitcoincharts.com/markets/, where the value of 1-Bitcoin, which was about 5 US dollars in September 2011, approximates 4,000 US dollars in August 2017.

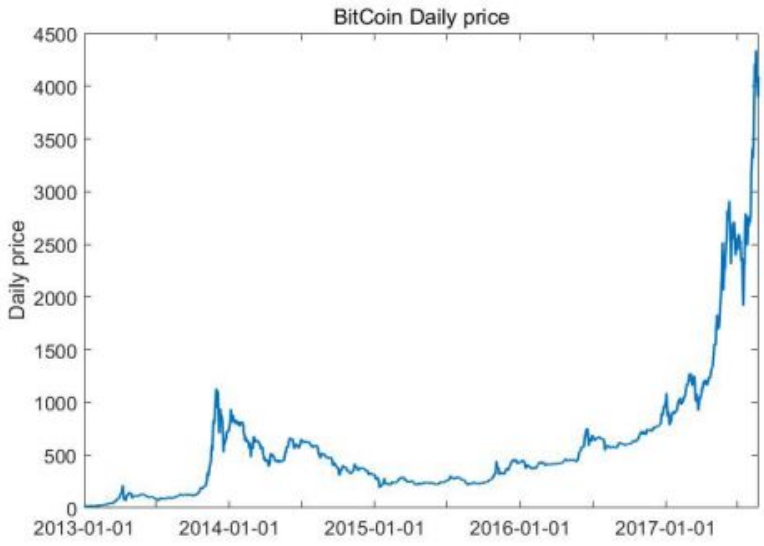

Fig. 3.1: Bitcoin daily price(USD), from Sep-11 2011 to Aug-22 2017

During this period (from Sep-11 2011 to Aug-22 2017), market volatility with enormous price changes in Bitcoin becomes exceptional compared with that in traditional currency markets.

The general price level of goods or services, P, can be determined indirectly from the global macroeconomic indexes in actual markets. The exchange rate between several fiat currencies and Bitcoin price describes the relationship between actual markets and Bitcoin market.

The main difference between the Bitcoin market and general currency markets originates from the fact that the Bitcoin is a virtual currency based on Blockchain technologies. Therefore, economic size, E; the velocity, V; and the capacity of the Bitcoin market, B, are closely related with several measurable market variables extracted from the Blockchain platform.

3.2 Blockchain

The Bitcoin blockchain is a public ledger that records bitcoin transactions. It is implemented as a chain of blocks, each block containing a hash of the previous block up to the genesis block of the chain and this leads up to the genesis block, which is the ever first block mined by Satoshi Nakamoto himself. A network of communicating nodes running bitcoin software maintains the blockchain. Transactions of the form payer X sends Y bitcoins to payee Z are broadcast to this network using readily available software applications.

Network nodes can validate transactions, add them to their copy of the ledger, and then broadcast these ledger additions to other nodes. To achieve independent verification of the chain of ownership each network node stores its own copy of the blockchain. About every 10 minutes, a new group of accepted transactions, called a block, is created, added to the blockchain, and quickly published to all nodes, without requiring central oversight. This allows bitcoin software to determine when a particular bitcoin was spent, which is needed to prevent double-spending.

By design, a blockchain is resistant to modification of the data. It is ”an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way”. For use as a distributed ledger, a blockchain is typically managed by a peer-to-peer network collectively adhering to a protocol for inter-node communication and validating new blocks.

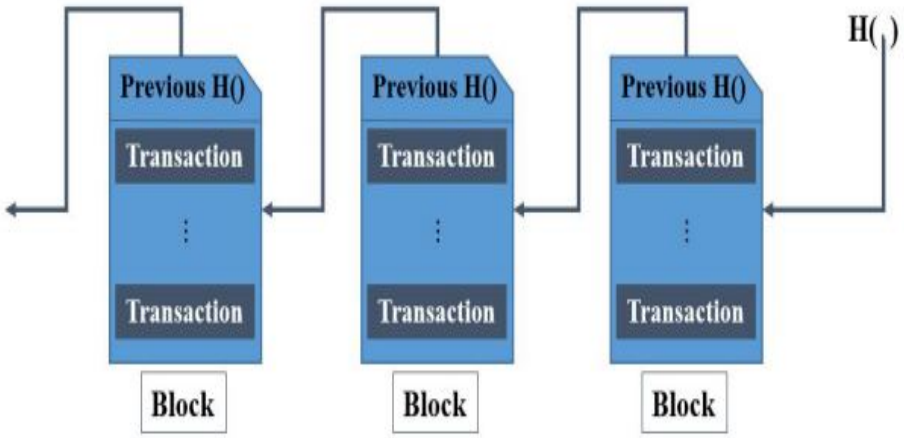

Fig. 3.2: Formation of Blockchain

Once recorded, the data in any given block cannot be altered retroactively without alteration of all subsequent blocks, which requires consensus of the network majority. Although blockchain records are not unalterable, blockchains may be considered secure by design and exemplify a distributed computing system with high Byzantine fault tolerance. Decentralized consensus has therefore been claimed with a blockchain. Figure 3.2 shows the general structure of Blockchain.

A participant in a Bitcoin network acts as a part of a network system by providing hardware resources of their own computer, which is called a distributed system. All issuance and transaction of money are conducted through P2P networks. All trading history is recorded in the Blockchain and shared by the network, and all past transaction history is verified by all network participants.

The unit called block, which includes recent transactions and a hash value from the previous block, creates irreversible data by a hash function, and is pointed out from the next block. It takes more than a certain amount of time to generate the block to make impossible to forge all or part of the Blockchain. This algorithm is called proof of work (PoW), and the difficulty is automatically set to ensure that the problem can be solved within approximately 10 minutes. PoW also provides incentives to motivate participants to maintain the value of Bitcoin

by paying Bitcoin for the participant who created the block.

Decentralization is the value pursued by all cryptocurrencies as opposed to general fiat currencies being valued by central banks. Decentralization can be specified by the following goals: (i) Who will maintain and manage the transaction ledger? (ii) Who will have the right to validate transactions? (iii) Who will create new Bitcoins? The blockchain is the only available technology that can simultaneously achieve these three goals. Generation of blocks in the Blockchain, which is directly involved in the creation and trading of Bitcoins, directly influence the supply and demand of Bitcoins [5]. Combination of Blockchain technologies and the Bitcoin market is a real-world example of a combination of high-level cryptography and market economies.

3.2.1 Mining

Mining is a record-keeping service done through the use of computer processing power. Miners keep the blockchain consistent, complete, and unalterable by repeatedly grouping newly broadcast transactions into a block, which is then broadcast to the network and verified by recipient nodes. Each block contains a SHA-256 cryptographic hash of the previous block, thus linking it to the previous block and giving the blockchain its name.

To be accepted by the rest of the network, a new block must contain a so-called proof-of-work (PoW).The PoW requires miners to find a number called a nonce, such that when the block content is hashed along with the nonce, the result is numerically smaller than the network's difficulty target. This proof is easy for any node in the network to verify, but extremely time-consuming to generate.

Every $2,016$ blocks (approximately 14 days at roughly 10 min per block), the difficulty target is adjusted based on the network's recent performance, with the aim of keeping the average time between new blocks at ten minutes. In this way the system automatically adapts to the total amount of mining power on the network.The proof-of-work system, alongside the chaining of blocks, makes modifications of the blockchain extremely hard, as an attacker must modify all subsequent blocks in order for the modifications of one block to be accepted. As new blocks are mined all the time, the difficulty of modifying a block increases as time passes

and the number of subsequent blocks increases.

3.2.2 Proof of work

A proof of work is a piece of data which is difficult (costly, time-consuming) to produce but easy for others to verify and which satisfies certain requirements. Producing a proof of work can be a random process with low probability so that a lot of trial and error is required on average before a valid proof of work is generated. Bitcoin uses the Hashcash proof of work system.

Proof of work makes it extremely difficult to alter any aspect of the blockchain, since such an alteration would require re-mining all subsequent blocks.. It also makes it difficult for a user or pool of users to monopolize the network's computing power, since the machinery and power required to complete the hash functions are expensive.

PoW agreement algorithm comes with several inherent risks.

- First, the validity of the block can be intervened when the majority of total participants is occupied by a group with a specific purpose called 51 percentage problem.
- Second, when the Blockchain is forked, a considerable amount of time is consumed to form the agreed Blockchain until the longest chain is selected after generation of several blocks. This condition causes a transaction delay because the transaction cannot be completed during that time.
- Lastly, there may be the capacity limit of the Blockchain or the performance limit of each node.

3.2.3 Applicaions of blockchain

Cryptocurrency economy is by now the most popular application of blockchain technology and also the most controversial one since it enables a multibillion-dollar global trading market of essentially anonymous transactions without government control. Compared to previous digital cash constructions, the blockchain based ones ingeniously combine the distributed consensus protocol, point-to-point communication and PoW (Bitcoin) techniques to

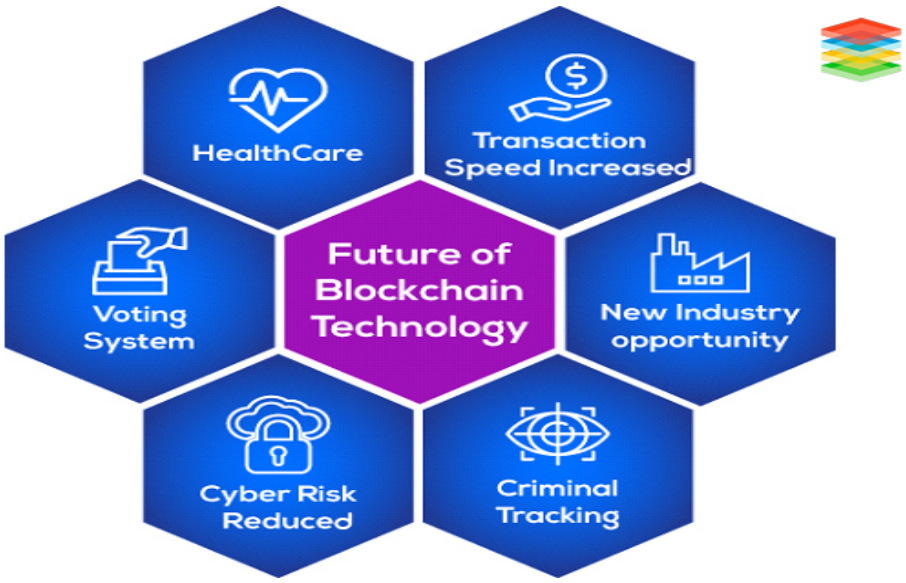

Fig. 3.3: Future of BLockchain Technology

prevent double-spend attacks and remove the need for a trusted party. Fig.3.3 shows the applications of Blockchain other than cryptocurrency.

Smart contracts are contracts which are automatically enforced by computer protocols featuring the same kind of agreement to act or not act without the need for trust between parties. Smart contracts were first proposed by Szabo in 1996 and with blockchain, which can be regarded as a distributed state machine without trusted third parties, can now be brought into reality. Although the functionality is limited due to a small instruction set that is not Turingcomplete, Bitcoin do support a small set of smart contracts. Later on, the most notable open source project Ethereum aims at providing a Turing-complete programming language to support arbitrary code execution on its blockchain, which in turn supports any kind of smart contracts.

Smart property: A tangible or intangible property, such as cars, houses, or cookers, on the one hand, or patents, property titles, or company shares, on the other, can have smart technology embedded in them. Such registration can be stored on the ledger along with contractual details of others who are allowed ownership in this property. Smart keys could be used to facilitate access to the permitted party. The ledger stores and allows the exchange of these smart keys once the contract is verified.The decentralized ledger also becomes a system for

recording and managing property rights as well as enabling the smart contracts to be duplicated if records or the smart key is lost.Making property smart decreases your risks of running into fraud, mediation fees, and questionable business situations.

Payments: cross-border payments: The global payments sector is error-prone, costly, and open to money laundering. It takes days if not longer for money to cross the world. The blockchain is already providing solutions with remittance companies such as Abra, Align Commerce and Bitspark that offer end-to-end blockchain powered remittance services. In 2004, Santander became one of the first banks to merge blockchain to a payments app, enabling customers to make international payments 24 hours a day, while clearing the next day.

3.3 Bayesian neural networks

Bayesian neural networks (BNN) is a transformed Multilayer perceptron (MLP) which is a general term for ANNs in the fields of machine learning. The networks have been successful in many application such as image recognition, pattern recognition, natural language processing, and financial time series.

3.3.1 Artificial neural network

An artificial neuron network (ANN) is a computational model based on the structure and functions of biological neural networks. Information that flows through the network affects the structure of the ANN because a neural network changes - or learns, in a sense - based on that input and output.ANNs are considered nonlinear statistical data modeling tools where the complex relationships between inputs and outputs are modeled or patterns are found.ANN is also known as a neural network.

ANNs are composed of multiple nodes, which imitate biological neurons of human brain. The neurons are connected by links and they interact with each other. The nodes can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. The output at each node is called its activation or node value.Each link is associated with weight. ANNs are capable of learning, which takes place by altering weight values.

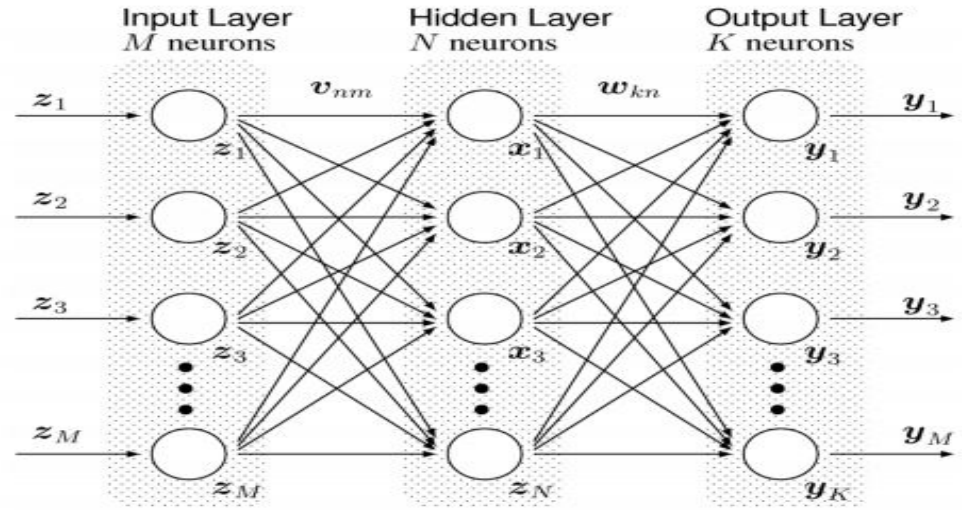

Fig. 3.4: Schematic view of a Neural Network.

The Fig.3.4 depicts the general outline of an artificial neural network. There are two artificial neural network topologies FeedForward and Feedback.

1. FeedForward: In this ANN, the information flow is unidirectional. A unit sends information to other unit from which it does not receive any information. There are no feedback loops. They are used in pattern generation/recognition/classification. They have fixed inputs and outputs.

2. Feedback: Here, feedback loops are allowed. They are used in content addressable memories.

3.3.2 Bayesian network

These are the graphical structures used to represent the probabilistic relationship among a set of random variables. Bayesian networks are also called Belief Networks or Bayes Nets. BNs reason about uncertain domain.

In these networks, each node represents a random variable with specific propositions. For example, in a medical diagnosis domain, the node Cancer represents the proposition that a patient has cancer.The edges connecting the nodes represent probabilistic dependencies among

those random variables. If out of two nodes, one is affecting the other then they must be directly connected in the directions of the effect. The strength of the relationship between variables is quantified by the probability associated with each node.

The structure of a BNN is constructed with a number of processing units classified into three categories: an input layer, an output layer, and one or more hidden layers.

Bayes rule : In probability theory and statistics, Bayes' theorem (Bayes' law or Bayes' rule) describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

Weights of a BNN must be learned between the input-hidden layer and hidden-output layer. Backpropagation refers to the process in which weights of hidden layers are adjusted by the error of hidden layers propagated by the error of the output layer. An optimization method called delta rule is used to minimize the difference between a target value and output value when deriving backpropagation algorithm. In general, BNNs minimize the sum of the following errors, E_B, using backpropagation algorithm and delta rule.

$$E_B = \frac{\alpha}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} (t_{nk} - o_{nk})^2 \qquad \text{(Equ:3.4)}$$

where E_B is the sum of the errors, N is the number of the training variables, K is the size of the output layer, t_{nk} is the k-th variable of the n-th target vector, o_{nk} is the k-th output variable of the n-th training vector, α and β are the hyperparameter, and w_B is the weights vector of the Bayesian neural network

3.4 Resampling

Resampling is the method that consists of drawing repeated samples from the original data samples. The method of Resampling is a nonparametric method of statistical inference. Resampling involves the selection of randomized cases with replacement from the original data sample in such a manner that each number of the sample drawn has a number of cases that are similar to the original data sample. Due to replacement, the drawn number of samples that are

used by the method of resampling consists of repetitive cases.

Resampling generates a unique sampling distribution on the basis of the actual data. The method of resampling uses experimental methods, rather than analytical methods, to generate the unique sampling distribution. The method of resampling yields unbiased estimates as it is based on the unbiased samples of all the possible results of the data studied by the researcher[6].

Assumptions:

- This method of resampling generally ignores the parametric assumptions that are about ignoring the nature of the underlying data distribution. Therefore, the method is based on nonparametric assumptions.
- In resampling, there is no specific sample size requirement. Therefore, the larger the sample, the more reliable the confidence intervals generated by the method of resampling.
- There is an increased danger of over fitting noise in the data. This type of problem can be solved easily by combining the method of resampling with the process of cross-validation.

3.4.1 Bootstrap resampling

A bootstrap method is one of the sampling techniques that new data set is sampled from the original data set with the replacement. Bootstrapping is any test or metric that relies on random sampling with replacement. Bootstrapping allows assigning measures of accuracy to sample estimates. This technique allows estimation of the sampling distribution of almost any statistic using random sampling methods.
A typical bootstrap works as follows:

1. We have the original data set D with the number of N.
2. Below following step is repeated B times for particular large number to produce B different bootstrap data set, $Z_1, Z_2, ..., Z_B$. Data set Z_i with the size $N$is generated by sampling from the original data set D with the replacement.
3. The machine is trained from each bootstrap data set.

4. Accuracy of the machine is calculated by averaging each bootstrap data set.

$$Accuracy = \frac{1}{B}\sum_{j=1}^{B}\frac{1}{N}\sum_{i=1}^{N}(1 - Loss(\hat{y}_i^j, y_i)) \quad \text{(Equ:3.5)}$$

where y_i is an i-th true training output data, $\hat{y}_i^j$ is an i-th estimated output from the bootstrap data Z_j , and $Loss(.,.)$ is a loss function.

3.4.2 Cross-validation resampling

A cross-validation randomly divides the original data set into K equal-sized parts without the replacement. We fit the machine learning model to the $K - 1$ parts leaving out particular set k and acquire a prediction error for the left-out k part. Total prediction accuracy is combined after the procedure is repeated for each part to leave. A general procedure is as follows:

1. We divide the original data set into K partial equal-sized data set,$C_1, C_2, ..., C_K$, without the replacement. nk is the number of each partial set defined by n/K.

2. We can compute the total accuracy:

$$Accuracy_K = \sum_{k=1}^{K}\frac{n_k}{N}\frac{1}{n_k}\sum_{i=1}^{n_k}(1 - Loss(\hat{y}_i^k, y_i)) \quad \text{(Equ:3.6)}$$

where N is the total number of the original data set, others have same definition with in the bootstrap description.

3. The estimated standard deviation of the cross-validation:

$$SE(CV_k) = \sqrt{\frac{\sum_{k=1}^{K}(Err_k - Err_k)^2}{N-1}} \quad \text{(Equ:3.7)}$$

where Err_k is the k-th loss, $\sum_{i=1}^{K}$ Loss$(\hat{y}_i^k, y_i)$.

Types of cross-validation methods:

- k-fold cross validation : In K-fold cross-validation, the dataset X is divided randomly into K equal-sized parts, X_i , $i = 1, ..., K$. To generate each pair, we keep one of the K parts out as the validation set V_i, and combine the remaining $K1$ parts to form the training set T_i. Doing this K times, each time leaving out another one of the K parts out, we get K pairs (V_i, T_i):

$$V_1 = X_1, T_1 = X_2 U X_3 U ... U X_K \quad \text{(Equ:3.8)}$$

$$V_2 = X_2, T_2 = X_1 U X_3 U ... U X_K \quad \text{(Equ:3.9)}$$

$$V_K = X_K, T_K = X_1 U X_2 U ... U X_K 1 \quad \text{(Equ:3.10)}$$

- $5X2$ cross-validation: In this method, the dataset X is divided into two equal parts $X_1^{(1)}$ and $X_1^{(2)}$. We take $X_1^{(1)}$ as the training set and $X_1^{(2)}$ as the validation set. We then swap the two sets and take $X_1^{(2)}$ as the training set and $X_1^{(1)}$as the validation set. This is the first fold. the process id repeated four more times to get ten pairs of training sets and validation sets.

Comparison of bootstrapping and cross validation

Bootstrap is adequate to validate a predictive model performance, to use an ensemble method, and to estimate of bias and variance of the trained model. Bootstrap creating the cloned multiple samples with the replacement is not originally developed for model validation. It can give more biased results. Therefore, we employ the cross-validation technique to our model validation.

Cross-validation can create high-variance problems when data size is small [7]. Our data size is sufficient to overcome the problem. We employ the 10-fold cross-validation methods generally used for model validations.

3.5 Prediction of bitcoin price

3.5.1 Data specification

The standard economic theories are insufficient to account for the impressive price development and volatility of Bitcoin . Bitcoin markets do not possess purchasing power nor interest rate parity. This fact suggests that the need for completely new determinants of Bitcoin price: the Blockchain information that includes relevant features as main determinants for pricing Bitcoin.

- **Response variables**
  - Prices or log prices of Bitcoin (USD)
  - Volatility or log volatility of Bitcoin (USD)
- **Blockchain information**
  - Average block size : the size of a block verified by all participants.
  - Transactions per block : average number of transactions per block.
  - Median confirmation time : the median time for each transaction to be accepted into a mined block and recorded to the ledger.
  - Hash rate : estimated number of Tera (trillion) hashes per a second all miners (market participants to solve a hash problem for making a block) is performing.
  - Difficulty : next difficulty =(previous difficulty 2016 10 minutes)/(time to mine last 2016 blocks).
  - Cost percentage of a transaction : miners revenue as the percentage of the transaction volume.

  - Miners revenue : Total value of coin-base block rewards and transaction fees paid to miners.
  - Confirmed transaction : the number of confirmed the validity of transactions per day.
  - Total number of a unique Bitcoin : market capitalization of Bitcoin.

- **Macro economic devolopemental indices**
  - $SP500$, Eurostoxx, DOW30, NASDAQ,
  - Crude oil, SSE, Gold, VIX, Nikkei225, and FTSE100
- **Global currency ratio**
  - Exchange rates between major fiat currencies (GBP, JPY, CHF, CNY, EUR)

The above mentioned data are taken as the input for predicting the price and volatility of Bitcoin. The prediction of Bitcoin price is determined empirically using Linear Regression, Support Vector Regression and Bayesian Neural Network. This section includes the results obtained from the comparative study.

Time series modelling

A time series is a series of data points indexed in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data.

Time series analysis is a statistical technique that deals with time series data, or trend analysis. Time series data means that data is in a series of particular time periods or intervals. The data is considered in three types:

- Time series data: A set of observations on the values that a variable takes at different times.

- Cross-sectional data: Data of one or more variables, collected at the same point in time.
- Pooled data: A combination of time series data and cross-sectional data.

For time series analysis, nonlinear methods, such as kernel regression model, exponential autoregressive models, artificial neural network (ANN), BNN, and support vector regression [8], have attracted research interest and exhibited improved predictive performance for various time series data.Unlike other widely studied time series researches, there are few related papers analyzing the Bitcoin processes in terms of prediction performance. In this work, we have employed Bayesian neural networks since the predicted model with a large number.

3.5.2 Linear regression

Here a linear model for analysis of Bitcoin price is constructed. It follows the following assumptions:

- The model assumption that linear relationships exist between response variables and independent variables
- the residual assumption that Residual terms are independently and identically distributed.
- Elimination of several explanatory variables with large VIF values, to avoid multicollinearity problem.

Multicollinearity Problem Multicollinearity occurs when independent variables in a regression model are correlated. This correlation is a problem because independent variables should be independent. If the degree of correlation between variables is high enough, it can cause problems when you fit the model and interpret the results. Multicollinearity causes the following two basic types of problems:

- The coefficient estimates can swing wildly based on which other independent variables are in the model. The coefficients become very sensitive to small changes in the model.
- Multicollinearity reduces the precision of the estimate coefficients, which weakens the statistical power of your regression model.

16 suitable discriminators are selected after eliminating variables with large VIFs and perform linear regression analysis on Bitcoin log prices and log volatilities with these 16 discriminators.From these 16 regressors, we construct two linear models, one for the log price and one for the volatility of Bitcoin process. Finally, we generate histograms residuals of each model to verify the residual assumption by confirming it follows a normal distribution.

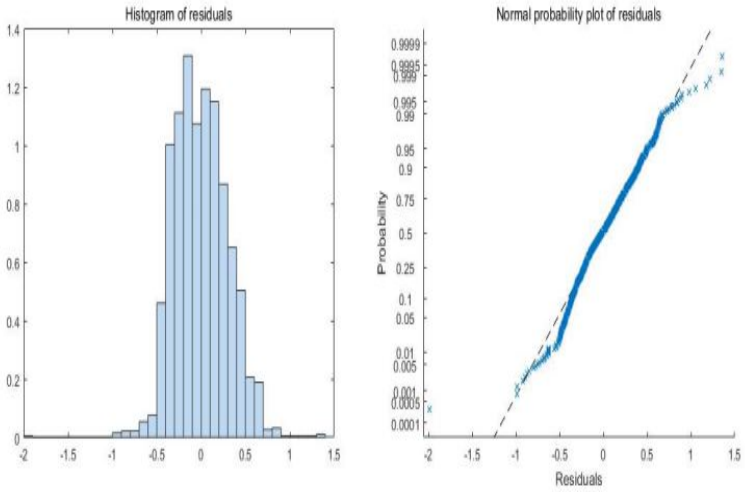

Fig. 3.5: Residual evaluations for Histogram, Normal probability (QQ) plot of the Bitcoin log price

From the figure 3.5, it is clear that the Bitcoin log price satisfies the residual assumption for linear regression: the histogram is bell-typed and symmetric and the QQ-plot shows a similar pattern with the normal distribution. By contrast, Figure 3.6 show that residuals of the linear model for log volatility of Bitcoin do not follow a normal distribution with a positive-skewed histogram.

A histogram that is bell-typed and symmetric, and the QQ plot showing a similar pattern with the normal distribution is said to satisfy the residual assumption of linear regression. On the other hand, a histogram and QQ plot which does not follow these constraints, is said to not have residual assumption property satisfied.

In regression analysis, the difference between the observed value of the dependent variable(y) and the predicted value $(\hat{y})$ is called the residual (e). Each data point has one residual.

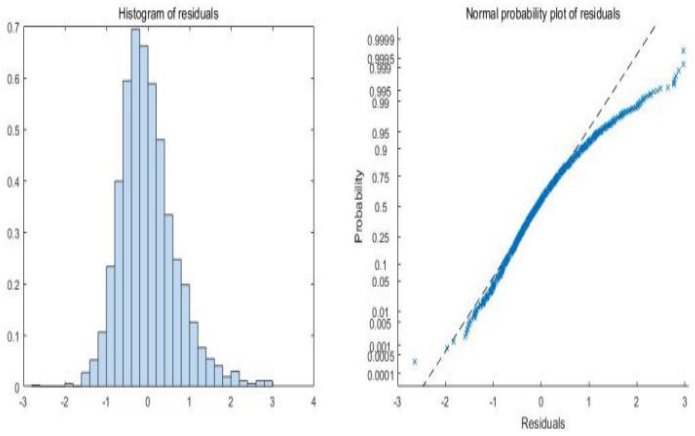

Fig. 3.6: Histogram, Normal probability (QQ) plot of the Bitcoin log volatility.

Residual = Observed value - Predicted value.

$$e = y - \hat{y}. \tag{Equ:3.11}$$

QQ plot is a plot showing residuals along x-axis and their probabilities along y-axis.

From these plots, we can conclude that the learned linear model does not make an adequate prediction of the output value albeit in predicting trends in little.

3.5.3 Support vector regression

Support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

When data is unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The support vector clustering algorithm, applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data, and is one of the most widely used clustering algorithms in industrial applications.

3.5.4 Bayesian neural networks

Why bayesian neural networks?

A Bayesian neural network (BNN) can naturally deal with increasing number of relevant features in the evaluation. A BNN includes a regularization term into the objective function to prevent the overfitting problem that can be crucial to our framework. When the machine considers a lot of input variables, a trained machine can be complex and suffer from the overfitting problem. BNN models showed their effect to the financial derivative securities analysis.

A BNN is a non-linear version of ridge regression, which is largely based on the Bayesian theory for neural networks. Unlike conventional neural networks that maximize marginal likelihood, BNN is a machine maximizing the value of posterior through an application of the Bayes theory. The elements added to the error term cause the machine to learn by selecting a weight with high importance even when the number of total weights is reduced rather than distributed to a large number of weights.

A total of 25 explanatory variables belonging to three categories are employed as inputs for BNN learning.We also address another input set that comprises 16 input variables by eliminating several unimportant variables as mentioned in the previous subsection.

Two response variables are considered,

- log price of Bitcoin
- volatility of Bitcoin price

We consider volatility of Bitcoin price, because extremely high volatility is an important feature of Bitcoin. We use logscaled values of both output response variables to account for the large difference between Bitcoin value in the early period and its most recent value. We train the BNN model through 10-fold cross-validation. To mitigate the effect of how to divide the data [9], we repeated hold-out validation steps where $\frac{9}{10}N$ training data and $\frac{1}{10}N$ test data, given the total number of the data is N. Where performances of each trained model are measured by root mean square error (RMSE) and mean absolute percentage error (MAPE). Definitions of each evaluation criteria are as followings:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \qquad \text{(Equ:3.12)}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} \qquad \text{(Equ:3.13)}$$

where N is the number of samples,y_i is the i-th true objective value, and $\hat{y}_i$ is the i-th estimated value.

Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are.

The Mean Absolute Percentage Error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of prediction accuracy of a forecasting method in statistics [10].

BNN models outperform other models in terms of RMSE and MAPE for predicting the

log price of Bitcoin. Log price of Bitcoin is learned exceptionally by the BNN model with training and test error of around 1 percent MAPE. In the case of log volatility, the prediction error of log volatility in the test phase is slightly larger than that in the training phase.

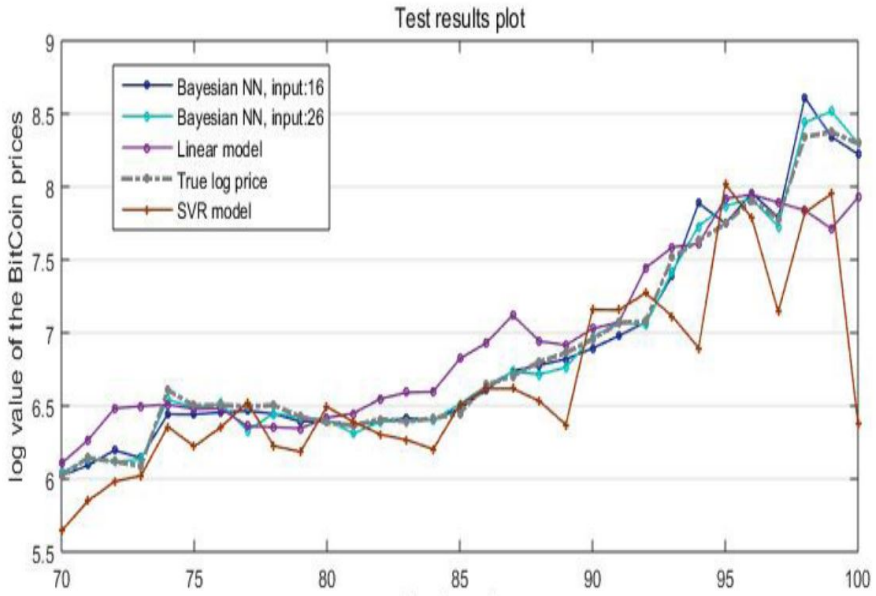

Fig. 3.7: Perfromance Evaluation of Bitcoin log Price

Figure 3.7 shows the values of estimated response variables for the recent 30 test input data according to time indexes. BNN model is more reliable for describing the process of log volatility than other benchmark models. After eliminating redundant variables from linear correlation analysis, the error value is relatively small when all 26 input variables are considered instead of the abridged 16 input variables. This condition implies that removed variables may explain nonlinear relationships to adequately account for response variables.

The recent volatile tendency was well expressed in terms of explanatory input variables. The case of log price presents a tendency for underestimation when price rises and overestimation when the price falls. In the case of the log price, we can see that all models predict the actual tendency of the price to some extent.

On the other hands, in terms of error size, it is confirmed that other models are larger than that of Bayesian neural networks. There is no tendency of over or under-estimate in all models. Bayesian neural networks tended to predict consistent trends regardless of the number of inputs.

In the case of volatility, the Bayesian NN model predicts better the direction of volatility than other benchmark models, and neither of the four models tends to over or underestimate.

Conclusion

Bitcoin is a successful cryptocurrency, and it has been extensively studied in fields of economics and computer science. The time series of Bitcoin price is analysed with a BNN using Blockchain information in addition to macroeconomic variables and the recent high volatile nature of Bitcoin prices are addressed.Through the empirical analysis, the BNN model describes the fluctuation of Bitcoin up to August 2017, which is relatively recent. Unlike other benchmark models that fail directional prediction, the BNN model succeeded in relatively accurate direction prediction. From these experimental results, the BNN model is expected to have similar performance in more recent data. As the variation of Bitcoin process gets attention, it is expected that the expansion and application of the BNN model would be effective for the analysis and prediction of the Bitcoin process.

References

[1] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, Tech. Rep., 2008

[2] A. H. Dyhrberg, Bitcoin, gold and the dollarA GARCH volatility analysis, Finance Res. Lett., vol. 16, pp. 8592, Feb. 2016.

[3] P. Katsiampa, Volatility estimation for Bitcoin: A comparison of GARCH models, Econ. Lett., vol. 158, pp. 36, Sep. 2017.

[4] A. F. Bariviera, M. J. Basgall, and W. Hasperu, and M. Naiouf, Some stylized facts of the Bitcoin market, Phys. A, Stat. Mech. Appl., vol. 484, pp. 8290, Oct. 2017.

[5] J. Chu, S. Nadarajah, and S. Chan, Statistical analysis of the exchange rate of Bitcoin, PLoS ONE, vol. 10, no. 7, p. e0133678, 2015.

[6] A. Urquhart, The inefficiency of Bitcoin, Econ. Lett., vol. 148, pp. 8082, Nov. 2016.

[7] S. Nadarajah and J. Chu, On the inefficiency of Bitcoin, Econ. Lett., vol. 150, pp. 69, Jan. 2017.

[8] A. H. Dyhrberg, Hedging capabilities of Bitcoin. Is it the virtual gold? Finance Res. Lett., vol. 16, pp. 139144, Feb. 2016.

[9] E. Bouri, P. Molnr, G. Azzi, D. Roubaud, and L. I. Hagfors, On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? Finance Res. Lett., vol. 20, pp. 192198, Feb. 2017.

[10] E.-T. Cheah and J. Fry, Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin, Econ. Lett., vol. 130, pp. 3236, May 2015.