FORECASTING BITCOIN PRICE VOLATILITY
WITH THE HYBRID ARMA-GARCH MODEL

A THESIS
Presented to
The Faculty of the Department of Economics and Business
The Colorado College

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Arts

By
John Ye
May 2018
Forecasting Bitcoin Price Volatility with the Hybrid ARMA-GARCH Model

John Ye

March 2018

Mathematical Economics

Abstract

This paper uses Bitcoin hacking and Bitcoin miners’ distribution to gain a deeper understanding of the Bitcoin price volatility. This paper uses data directly from the Bitcoin blockchain, as well as data gathered from various Internet sources. A two-step ARMA-GARCH model is chosen to detrend and interpret the Bitcoin price volatility. With three variables, the model can accurately forecast 5 percent of the total volatility. From the result, Bitcoin hack, scam and theft events would make the Bitcoin price more volatile, and more concentrated Bitcoin miners in the network would decrease the Bitcoin price volatility. However, because of the small percent of forecasted values, most of the Bitcoin price volatilities remains a mystery.

Key Words: Bitcoin, Cryptocurrency, Blockchain, ARMA, GARCH, Price Volatility

JEL Codes: D01
Dedication

For my advisor Dr. Daniel K. N. Johnson and all my friends, thank you for your encouragement and support.
ON MY HONOR, I HAVE NEITHER GIVEN NOR RECEIVED
UNAUTHORIZED AID ON THIS THESIS

John Ye
# TABLE OF CONTENTS

**ABSTRACT**

1. INTRODUCTION............................................................................................................ 1
2. LITERATURE REVIEW........................................................................................................ 4
3. DATA.................................................................................................................................... 8
   a. Dependent Variable-Bitcoin Price Volatility................................................................. 9
   b. Independent Variable-Hacking Event........................................................................... 10
   c. Independent Variable-Miners’ Concentration............................................................ 11
4. MODEL................................................................................................................................ 14
   a. ARMA.......................................................................................................................... 14
   b. ARCH............................................................................................................................ 15
   c. ARMA-GARCH............................................................................................................ 17
   d. Information Criterions................................................................................................. 17
5. ANALYSIS............................................................................................................................ 19
6. CONCLUSIONS AND IMPLICATIONS............................................................................. 25
   a. Non-Seasonality effect on Bitcoin price volatility....................................................... 25
   b. The relationship between hacking and Bitcoin price volatility............................... 25
   c. Miner centralization to Bitcoin price volatility......................................................... 26
7. WORK CITED....................................................................................................................... 28
8. WORK CONSULTED......................................................................................................... 34
List of Tables

1. Summary of Independent Variables................................................................. 9
2. Summary of Hacking Independent Variables.................................................. 11
3. Summary of Herfindahl Independent Variables............................................. 13
4. ARMA Model Information Criteria................................................................. 20
5. ARMA Results.................................................................................................. 20
6. GARCH Results............................................................................................... 23
List of Figures

1. Hashrate Distribution on February 12th, 2018 ................................................................. 2
2. Independent Variable Line Graph .................................................................................... 10
3. Correlogram of Dependent Variable .............................................................................. 19
4. Partial Correlogram of Dependent Variable .................................................................... 20
5. Two-way Line Graph (Original Volatility, ARMA Forecast) ........................................ 22
**Introduction**

Bitcoin and other cryptocurrencies have attracted an increasing volume of media exposure in 2017 (Jones, 2018), which is both reflected by the surge in public interest and prices (Gannon, 2017). In 2017, Bitcoin price has skyrocketed from less than $1,000 in January to $18,000 in December, resulted a gain of 1338%. This is a dramatic price change for a currency with over $200 billion market cap and 14 million active users (Blockchain.info, 2018). Cryptocurrency related businesses have grown to over 240 companies across 28 industries; investments for cryptocurrency related companies has exceeded $2.12 billion based on PwC DeNovo insights. Interest in Bitcoin has also reached an all-time high on December 17th and became the second most searched term in global news, according to Google 2017 annual report. The volatility of Bitcoin price tightly influences many sectors in the finance industry. Its high volatility, as well as expected high volatility in the short future, has been considered as one of the main reasons why Bitcoin cannot grow to replace centralized government-issued currencies such as U.S. Dollars (Lambert, 2017). Shortly after the $18,000 peak price, Bitcoin price in U.S. Dollars dropped sharply in early February 2018 and reached a staggering $6,000. As a currency, Bitcoin is perhaps the most volatile of its kind (Williams-Grut, 2018).

Accompanying the dramatic volatility in Bitcoin price is the centralization in Bitcoin mining. Bitcoin mining is a service that provides Bitcoin creation and transaction bookkeeping service at the cost of CPU processing power. Before 2011, all Bitcoins were mined by anonymous miners across the globe. However, by 2018, this has dramatically changed. As the difficulty of mining a bitcoin increases over time, corporations and pooled miners make up the majority of Bitcoin mining. To better measure the speed of Bitcoin mining, hash rate was invented as a standardized unit for all miners. According to Blockchain.info, one of the largest Bitcoin API providers in the world, the top four Bitcoin mining pools dominate 69% of the entire
hash rate in the Bitcoin network, as of February 2018 (Figure 1). This figure is powerful evidence of how concentrated Bitcoin market has become today.

![Hashrate Distribution on February 12th, 2018](image)

**Figure 1: Hashrate Distribution on February 12th, 2018**

Bitcoin is also most popularly associated with hacking and other illegal activities because of its anonymous and circulating power (Cheng, 2017; Sulleyman, 2018). In April 2017, one of the largest global ransomware, WannaCry, along with its mutants, infected more than 200,000 computers across 150 countries (Shea, 2017). Many large organizations in both civil and government sectors were infected; some include major automobile manufactory, telecommunication providers, national universities, and state governments. The only form of payment demanded by the unidentifiable hacker was Bitcoin. The economic damage of this
cyber-attack was believed to exceed $4 billion, according to CBS News. The price of Bitcoin nearly doubled in a month following the event (Palmer, 2017).

Because of the anonymous and untraceable nature, Bitcoin scamming, hacking and stealing have been prevalent (Jefferies, 2013; Shen, 2018). In 2014, Mt. Gox, the largest cryptocurrency trader, reported a loss of 850,000 Bitcoin, worth around $473 million and 14% of total Bitcoin in circulation (McMillan, 2014; Norry, 2017; Kharif, 2018). Bitcoin price dropped 15% following the event (Blockchain, 2018). In August 2016, the second largest Bitcoin exchange platform, Bitfinex, lost 120,000 bitcoins in a security breach, which was valued at $72 million (Higgins, 2016). In December 2017, more than $70 million worth of Bitcoin was stolen from a Bitcoin mining company, NiceHash (Iyengar, 2017). News on Bitcoin hacking has been countless since the beginning of Bitcoin invention (Cheng, 2017).

The juxtaposed phenomena of Bitcoin price volatility and hacking news, as well as the trend of mining centralization, raises the question: is the volatility of bitcoin price influenced by hacking news and mining centralization? This paper will study the stochastic process of both mining centralization and bitcoin hacking to Bitcoin price volatility. This paper will attempt to establish a model to forecast the short-term volatility change based on news and market trend. With that approach, this paper will continue with literature review, data, model, analysis, and conclusion sections. To better understand Bitcoin price volatility, the literature review section will cover previous studies on this topic.
Literature Review

Bitcoin is designed to facilitate a transfer of value between parties. It uses cryptography to maintain the value of each note (Satoshi, 2008). Mainstream voices argue that Bitcoin is to be considered as an asset rather than a currency. Yermack (2013) found its price behaves with no correlation with any currency and should be treated as a speculative investment. Gronwald (2014) also found that Bitcoin’s extreme price movement can be observed in immature financial markets. In this context, this paper will consider Bitcoin as an asset for the analysis. For the rest of this section, I will review the previous Bitcoin research on the topics of volatility, shocking events, miner equilibrium, and econometric models. This section will motivate the topic by filling the gap of previous research and select the optimal model for this study.

Past research has questioned Bitcoin’s volatility. From 2012 to 2014, the daily variance of the Bitcoin USD exchange rate remained stable. For the same period, BTC USD exchange rate had increased 50-fold, and the biggest Bitcoin trading platform, Mt. Gox, had filed bankruptcy. The returns of Bitcoin clearly cannot be fitted with a single distribution (Chan et al. 2007). A study that specifically focused on the 2013 Bitcoin price crash revealed that before the 2013 crash, positive shocks increased the conditional volatility more than negative shocks, but after the 2013 crash, shocks and volatility had no significant relation (Bouri et al. 2017).

Some previous research has been able to attribute some Bitcoin price volatility to events besides the 2013 crash. Hencic and Gourieroux (2015) found episodes of local trends can be interpreted as speculative bubbles. They argue that online trading caused much of the bubbles. Baek and Elbeck (2013) detrended Bitcoin price volatility with S&P 500 trend ratio and compared those values. The result was insignificant. Thus the study concluded that the Bitcoin price volatility was a mostly speculative vehicle. Kristoufek (2013) found that Bitcoin price and...
Google Trends are highly correlated. They demonstrated that when the price is above or below the trend, a spike of interest will appear.

The similar effect of Bitcoin price volatility from hacking events can be traced in the history of U.S. dollars. During the suspension of convertibility from greenback to gold from 1862 through 1878, public expectation of government supply and potential depreciation had not caused changes in exchange rates and prices, but the expectation of resumption did (Calomiris, 1988). The study showed that exchange rate disturbance was directly correlated to the market efficiency at the time. Government debt was the most important influence on the forecast variance of the exchange rate. Bessembinder and Seguin (1993) found that unexpected events have larger impacts than expected on the price volatility.

Beyond Bitcoin hacking, other factor of interest is the distribution of miners. Miners are essential to the equilibrium of Bitcoin system, as Kroll et al. showed in 2013, miners build the strength of Bitcoin and receive rewards for their efforts; different behaviors of each miner in the Bitcoin system can lead to infinite equilibria. The reward system and the miners play a game that ensures that the Bitcoin system runs. However, this mechanism can only defend attacks from small mining companies. If incentives do not hold, the majority of miners can destroy the Bitcoin system. Eyal and Sirer (2014) presented their analysis that if colluding miners can obtain a greater reward than their fair share, minority will join the colluding party and become the majority. At this point, the Bitcoin system will cease to be a decentralized currency.

Decentralization is one of the key features of the Bitcoin. Several studies have focused on measuring this nature. Gencer et al. (2018) proposed a metric to study Bitcoin decentralization by measuring the network resources of nodes and the interconnection among them. Nodes are the validators, often individual computers, in the blockchain that verify transactions. This network resource measurement technique reflects the robustness of each node against attack. The study observed growth in Bitcoin major datacenter and the potential of centralization. However,
Internet speed is insufficient in measuring the actual mining speed, the computing power, and the computing power is how miners differ in the market share. For measurement of market concentration, I will be using Herfindahl index of the entire mining market for this analysis. Although other measurements for the market concentration exist such as Gini coefficient, Herfindahl Index is the most suitable because of the nature of this data (large sample size and time sensitive).

This paper will build upon existing research methods on price volatility. Moving Average is widely used as a means to evaluate asset volatility for risk management (Pesaran & Zaffaroni, 2005; Lux & Kaizoji, 2007). Autoregressive-moving-average model (ARMA), also known as the Box-Jenkins method, is popularized by Box and Jenkins in their 1970 book “Hypothesis testing in time series analysis.” On top of the traditional moving average approach of volatility modeling, the ARCH model has used autocovariance to represent volatility in a time series data. The ARCH (Autoregressive Conditional Heteroskedasticity) processes were first developed in 1982 by Robert Engle to capture the stochastic effects where a recent past in a series forecasts the variance of the next period.

However, because of the explosive effect of hacking, the analytical model cannot be simple ARMA or GARCH (Generalized ARCH). Nwogugu (2006) critiques the existing models of market risk (ARMA, GARCH, ARCH, EVT, VAT, Stochastic-Volatility, etc.). He argues that those models are inaccurate and inadequate in emerging markets and do not account for multifaceted risk and decision making. For this study, I will use a two-step hybrid ARMA-GARCH model for the analysis.

A hybrid ARMA-GARCH model has been used many times in previous financial volatility research. Jansky and Rippel found that forecasting with the ARMA-GARCH model is highly adaptive. Different from VAR (Vector Autoregression) forecasting models, ARMA-GARCH generates high accuracy for periods of low and high volatility. Albu et al. (2014) have
used the ARMA-GARCH model to estimate the impact of quantitative easing on credit risk. The model was able to forecast with impulsive exogenous events such as dramatic policy change. This paper will adopt the ARMA-GARCH approach to detrend and evaluate the explosive volatility effect from hacking as well as the mining concentration.

This paper aims to deepen the understanding of Bitcoin volatility by speculative influences from hacking and centralization. The analysis will use the ARMA-GARCH model to measure the effect of mining concentration and hacking activities.
Data

The raw data used for this paper are a series of historical figures obtained from internet sources. The Bitcoin daily USD price is daily tick data of the market recorded at 00:00 UTC, according to blockchain.info. Bitcoin price data are scraped from Bitfinex, the largest cryptocurrency trading platform since 2014. The Bitcoin hack, loss, and scam data are manually collected from major news sources as well as Bitcoin news sources such as bitcointalk.org and coindesk.com.

For the analysis, I chose the start date of December 2nd, 2011 and the end date of October 25th, 2017. The start date is shortly after the founding of Bitfinex and blockchain.info, two of the data providers, and this ensures the authenticity of the data from the providers. Furthermore, a start date of December 2011 was also deliberately chosen, because it was the beginning of market diversification. The data regarding miner identity is entirely unknown for dates before December 2011.

Bitcoin data may have been less accurate in the earlier days than later because the cryptocurrency market had not developed sophisticated market functions such as arbitraging (Badev & Chen, 2014), thus Bitcoin price was inconsistent across different platforms and different countries. Although this phenomenon still exists as a minor problem today, the Bitcoin price data are much more reflective of the market as a whole when the data are from a single source. Some other data used for this paper, such as miners’ distribution, may also be less accurate in the early days than later because of the small number of observations and the lack of technology of data collection.
Dependent Variable-Bitcoin Price Volatility

This paper aims to gain a better understanding of Bitcoin price volatility. To represent price volatility, I have taken the first order difference of daily bitcoin price as d.BTC USD Price in Table 1. Although Bitcoin price has been growing since the beginning of the sample observation, the nature of Bitcoin does not support natural or exponential growth. Thus, I have treated Bitcoin as a commodity and calculated the daily return rate of Bitcoin by dividing the first order difference with previous day’s price and obtained the daily realized return rate, which is labeled as d.BTC USD Price% in Table 1. Those data summarization are as follows:

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Observation</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTC USD Price</td>
<td>2,155</td>
<td>959.9799</td>
<td>922.1625</td>
<td>2.93</td>
<td>6020.37</td>
</tr>
<tr>
<td>d.BTC USD Price</td>
<td>2,154</td>
<td>2.444634</td>
<td>48.69882</td>
<td>-641.6499</td>
<td>633.4199</td>
</tr>
<tr>
<td>d.BTC USD Price%</td>
<td>2,154</td>
<td>0.0045022</td>
<td>0.0453423</td>
<td>-0.380167</td>
<td>0.431597</td>
</tr>
</tbody>
</table>

Table 1: Summary of Independent Variables

d.BTC USD Price% will be used as the dependent variable to study price volatility. From the simple line graph of Bitcoin daily rate of return, it is easy to observe that the mean is close to zero based on the symmetry at the 0 line. Also, the clustering effect can be visually observed as large volatilities happening adjacent to others. That evidence suggests that the dependent variable is weakly stationary. Furthermore, Durbin’s alternative test provides evidence of autocovariance, and the augmented Dickey-Fuller test reveals strong evidence against unit root (MacKinnon approximate p-value for Z(t)=0.0000).
From previous studies, Bitcoin hacking is speculated to be causing the volatility change in Bitcoin price. This paper will adapt Bitcoin hacking data in constructing the model. Each of the 60 events for bitcoin hack, loss, and scam are validated with at least two sources of information to ensure that the event influenced the public. Only events with more than 500 bitcoin losses were recorded to ensure that the magnitude of the news would have initiated market fluctuation. However, as Bitcoin price goes up and time goes on, this benchmark may not be sufficient. Market speculators may become accustomed to the quantity of the Bitcoin losses in the news or may react differently based on the Bitcoin price of the time. Neither the data or model can reflect this subjective sensation. Furthermore, Bitcoin hacking activities have become much more complicated over time. News about hacking on new cryptocurrencies such as Ripple and
Ethereum, as well as news about hacking with new methods, such as trojan mining viruses, have been distracting attention from the bitcoin hacking news in the old form. Market speculators’ reactions to the news have also evolved to be more complicated and thus difficult to quantify. Based on those considerations, the majority of the 60 events are between 2012 to 2013. Events between those two years are happened frequently and methods of hacking are relatively similar.

To better understand the hacking activities, I have separated the events into two components: the occurrence of the event (Bernoulli variable Hackoccure) and the amount of Bitcoin involved in each event (Hackamount). This paper will analyze those two variables separately and build an understanding of their results. Although it is better to detrend the number of Bitcoin stolen with Bitcoin price in U.S. dollar and U.S. dollar inflation rate across the years, the limited sample size does not need such modification. For dates without a record of activities, I have entered value 0 for both variables. The correlation between Hackoccure and Hackamount is 0.31, indicating low endogeneity between those two variables. The summary statistics for Bitcoin hacking is in the table 2.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Observation</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hackoccure</td>
<td>2,155</td>
<td>0.0176334</td>
<td>0.1316454</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hackamount</td>
<td>2,155</td>
<td>347.2854</td>
<td>6898.986</td>
<td>0</td>
<td>273209</td>
</tr>
</tbody>
</table>

Table 2: Summary of Hacking Dependent Variables

**Independent Variable-Miners’ Concentration**

Herfindahl-Hirschman Index, more commonly known as the Herfindahl index, is a measure of the size of individual firms about the industry. It delineates the competition structure in the market, such as the existence of monopolies or oligopolies, which subsequently measures the inequality in the market. Herfindahl index ranges between 0 and 10,000, the higher number
indicating a closer match with monopolistic competition. The calculation of Herfindahl index is simply the sum square of numerical values of the percent share, as follows:

\[
H = \sum_{i=1}^{N} s_i^2
\]  

(1)

where \(s_i\) is the numerical value of the market share of firm \(i\) in the market and \(N\) is the number of firms in the industry.

Calculating Herfindahl requires the total number of miners in the network. Determining the exact number is challenging because of the anonymous nature of the Bitcoin network. The number of miners is unknown, and no estimation of such has been done in previous research. The mining process requires using computer algorithms to solve puzzles, and the winner of the puzzle will be rewarded with one block to record transactions and to earn Bitcoin. This process heavily relies on the luck of the draw, thus the higher the computation power (hash rate), the higher the chance one can solve the puzzle before everybody else. As of 2018, one block is mined every 10 minutes in the Bitcoin network. The difficulty for finding a block has been consistently and continuously escalating since the beginning of the Blockchain. Chances for individual miners to find a new block gradually diminish. Miners, therefore, transition from individual miners with a single computer to mining companies and mining pools with multiple participants. (Eyal, 2014)

Because of the nature of Bitcoin mining, I have defined miners in the Bitcoin network as owners of mined blocks in the blockchain. By this definition, the maximum number of miners under this assumption each day is around 150, the same as the number of blocks mined daily. The daily mined block data is transparent in the Blockchain file, and it is organized by smartbit.com.au. The number is very consistent throughout the time with summary statistics as follows:
<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Observation</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herfindahl</td>
<td>2,155</td>
<td>1041.26</td>
<td>443.3672</td>
<td>219.3421</td>
<td>1884.38</td>
</tr>
</tbody>
</table>

Table 3: Summary of Herfindahl Independent Variables

Each block mined by miners can be digitally signed, and every miner with a public identity claims the ownership of each mined block to show the market share. This is the foundation of calculating the distribution of miners in the network. By labeling blocks mined in a certain period and organizing by miners’ identities, data collectors can have a precise figure on the percent distribution of public miners. For unknown miners in the network, I have assumed an even distribution of mined blocks amongst every miner. Under this assumption, every unknown block in the network belongs to independent and identical unknown miners. This reflects the nature of anonymous miners.

The measure of price volatility has not been sufficiently developed. Financial institutions and academia diverge in both the method and understanding of price volatility. No prior study was able to demonstrate the optimal time interval for volatility calculation. Because of the uncertain nature of Bitcoin mining, the Herfindahl Index is calculated at the monthly interval to give a more precise reflection of the mining speed. BTC.com, one of the major Bitcoin mining companies as of Jan. 2018, provides a history of distribution of the largest mining pools. Their data is used to calculate the Herfindahl Index for this paper.

This paper will establish an understanding of Bitcoin price volatility by building a forecasting model based on only those three independent variables. The model is expected to reflect the behavior of Bitcoin hacking and miners’ distribution on the Bitcoin price volatility. From previous literature, Bitcoin hacking event occurrence and the amount of Bitcoin involved are expected to increase the volatility of Bitcoin price, and more concentrated Miners’ distribution is expected to increase the Bitcoin price volatility over time.
Model

For the analysis, this study will adopt the hybrid ARMA-GARCH model for analyzing the Bitcoin price volatility. This section will introduce ARMA and GARCH processes separately. Although both models have been widely used for time series analysis as a standard practice, they differentiate in the ways of modeling the volatility. Lastly, this section will present the two-step process for the study.

ARMA

The ARMA (p, q) model assumes that the current value in a time series is constructed from the past by both the linear value and the error. The ARMA is divided into AR (Autoregressive) and MA (Moving Average) parts. The AR part involves regressing the time series data $X_t$ with a few lags (p) of its own. The AR(p) process can be noted as below:

$$X_t = c + \alpha_{t-1}X_{t-1} + \alpha_{t-2}X_{t-2} + \cdots + \alpha_{t-p}X_{t-p} + \epsilon_t$$ (2)

where $\alpha_i$ are the parameters of the estimation results from different lags and $\epsilon_t$ is ideally white noise. The MA part regresses the error term from the current time with a few lags (q). The MA (q) can be noted as below:

$$E_t = \mu + \beta_{t-1}\epsilon_{t-1} + \beta_{t-2}\epsilon_{t-2} + \cdots + \beta_{t-q}\epsilon_{t-q} + \epsilon_t$$ (3)

where $\mu$ is the expected error of $X_t$, often assumed to be zero. The combined ARMA (p, q) model is therefore:

$$X_t = c + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{i=1}^{q} \beta_i \epsilon_{t-i} + \epsilon_t$$ (4)

For this paper, I am using the ARMA model to capture the stochastic process by specifying a conditional mean, which is created by the lags of volatility from known
autocorrelation. It observes the change in the rate of return by using the deviation from the conditional mean. The ARMA model also captures the lagging nature of market behavior, that is, the speed of information dispensing as well as market synchronizing. Also, one of our independent variables, Bitcoin hacking activities, only occurs at specific days in a period. However, the effect of such events has a lasting impact on the market with unknown duration. The MA part of the model can detect the duration of the impact of Bitcoin price volatility and detrend the lasting impact from the event occurrence.

Furthermore, the ARMA model can also capture the seasonal effect of the Bitcoin market. Bitcoin sales and trading happen every day including weekends, whereas other commodities and currencies can only be traded from Monday to Friday. This discrepancy in availability may create a seasonality to the Bitcoin price volatility. If seasonality exists, we can observe a distinction of autocorrelation at lag 7 and its multiples.

The model specification $(p, q)$ will be visually inspected from Correlogram (ACF) for $q$ and Partial Correlogram (PACF) for $p$. Stata’s ACF produces a graph of autocorrelations based on Bartlett’s formula (1946) for MA $(q)$ process. For our time series analysis, I will specify to display lag values from 1 to 100 (lag values larger than 100 are highly improbable for daily tick data). This ACF graph will show the autocorrelation for 100 days of interval. Any outlier from the 95% confidence interval will be used as lag operators for $p$. PACF will find the autocorrelations with only specific lag values without taking shorter lags into account. Outliers from the 95% confidence interval will be used as a lag operator for $q$.

**ARCH**

The important characteristic of the ARCH regression models is that they capture both the clustered error variances and the change in variance by imposing a conditional mean.
GARCH (Generalized ARCH) (Bollerslev, 1986) aims to model the error variance change in a time series dataset. Previous research has successfully used GARCH to model Bitcoin volatility (Dyhrberg, 2016). Chu et al. in 2017 have applied different Information Criterion to test the fitting of a series of GARCH family models to the Bitcoin financial data. The result only shows a marginal difference between eight different GARCH family models. For simplicity, this paper will use the standard GARCH model and adopt the parameterization of (1,1) from Chu et al. as the GARCH operator. The GARCH model process can be specified as follows:

\[ X_t = \mu_t + \sigma_t Z_t \] (5)

where \( \mu_t \) denotes the conditional mean of the series, \( \sigma_t \) denotes a volatility process, and \( Z_t \) is the stochastic piece. The variance in GARCH (1,1) has the following property:

\[ \sigma_t^2 = \omega + \alpha_1 Z_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \] (6)

for \( \alpha_1 > 0, \beta_1 > 0, \text{and } \omega > 0 \). The main feature of this variance process is that it captures both the autocovariance feature and the influence from the last period. The strength of influence from the last period, or the significance of error clustering, is represented by \( \alpha_1 + \beta_1 \), and weak stationarity holds if \( \alpha_1 + \beta_1 < 1 \).

The GARCH model requires the dependent variable to be weakly stationary to control the autocovariance. It assumes that the dependent variable is a function of the past. The Conditional Heteroskedasticity allows autocovariances to be captured by assuming that the autocovariances caused the conditional heteroskedasticity. However, if we did not know that the autocovariance exists, the model would simply allow the heteroskedasticity. Thus, Engle’s Lagrange Multiplier test is usually required as preliminary testing to determine if such autocovariance exists.
ARMA-GARCH

This paper will use a hybrid two-step ARMA-GARCH model to analyze volatility concerning hacking events and Herfindahl Index. This model offers the best combination of expressing the random shock events as regressive disturbances. The Breusch-Godfrey test has shown significant lag in the error term thus shocking events such as Bitcoin hacking have a lasting effect on the price volatility. This cannot be accurately depicted by the ARCH model because ARCH will treat the duration of the impact as part of the price volatility. Residuals from the ARMA model can detrend the error term and extract the shocking event from the variable. Then GARCH will reduce the clustered volatility to a single volatile day and analyze the volatility with dependent variables.

Information Criteria

Because the ideal model must be suitable for both fitting the data and generating the data, two information criterions will be accessed to compare the models among different configurations of the ARMA processes. Akaike Information Criterion (AIC) was first developed by Hirotugu Akaike to estimate the quality of a model with a given set of data. AIC estimates a certain value for a set of data to compare with different models. It provides an estimate of the amount of information lost by a model in the process of generating the data. Its function is as the following:

\[ AIC = 2k - 2 \ln(\hat{L}) \]  

where \( k \) is the number of estimated parameters in the model, and \( \hat{L} \) is the maximum value of the likelihood function for the model. The function punishes overfitting for the variable. Thus, lower the AIC value, better the model is fitted for the variables. On top of AIC, Bayesian Information Criterion (BIC) will also be used to access the fit of models. BIC function is defined as the follows:

\[ BIC = \ln(n)k - 2\ln(\hat{L}) \]
where \( n \) is the number of observations in the sample. Compared to AIC, BIC’s penalty term is even stronger. The combination of two can find the model that best fits and generates the data.
Analysis

Before applying the time series model, the variable must be tested against unit root as the existence of unit root would nullify the autoregressive effect. Since in the data section, the Dickey-Fuller test has already revealed that unit root does not exist for the BTC price change percent (the dependent variable), the series can be directly fitted with ARMA model. The parameters of the ARMA model, AR order, and MA order, are usually determined by a number of outliers before the convergence in the correlogram. The correlogram (ACF) and the partial correlogram (PACF) of the dependent variable are in the figures below and in the next page.

The two graphs show that the series is strongly distinct from the white noise as some lags visibly strike out from the 95% confidence band. However, no convergence or repetition at any lag multiple is observed. Strong outliers at lags of multiples of 7 (7, 14, 21…) were expected because Bitcoin is traded every day throughout the week whereas other commodities and currencies can only be traded from Monday through Friday. If the lack of trading or market reaction existed, this seasonality effect would be noted as strikes at lag 7. It is safe to conclude that Bitcoin volatility does not have any weekly seasonality, or in other words, it is hard for Bitcoin to have “Black Fridays.”
Because the graphs cannot find the optimal parameter pair for the ARMA model fitting, the combination can be determined by fitting every pair of parameterization and finding the most favorable model with both the highest Z-Score and the lowest Information Criteria (IC) value. Because of the limited computation power, I have assumed that the lag effects for volatility in Bitcoin market last for no more than ten days. It is very uncommon for events to shock the market for longer than ten days, and, according to PACF, the number of lags reduces after certain periods. For the Moving Average, I have chosen to use three independent variables to determine the lagging error term: the Herfindahl index, Hacking Occurrence, and Hacking Amount.

In total, 100 ARMA models have been run and their pseudo $R^2$ have been compared based on fitted values. Among all, (1, 3), (3, 1), (7, 9), (8, 9) showed the most significance, which means that their ARMA processes performed the most portion of the original structure. On top of that, I have also chosen a list of individual AR and MA terms based on the outliers from the DCF and PACF graphs: (1 3 5 6 10 15 16 24, 1 2 3 5 6 10 15 24). The AIC and BIC values of each model are listed in the table below.
Based on both the AIC and BIC results, the list of AR and MA term in Model 5 has been chosen as the parameterization for the ARMA model. Although it is not the best in both measures, the difference in AIC significantly outweighs the difference in BIC, compared to the second-best model (Model 2). BIC punishes over parameterization more than AIC, which is caused by a large number of AR and MA terms in model 5. Furthermore, from the information criterion result, it is noteworthy that the distinct lags in the higher term are not identical to white noise because model 3 and 4 yielded a lower AIC than model 1 and 2, which supports the conclusion from the correlogram.

Running the ARMA model 5, the results are as follows:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Z score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hackoccure</td>
<td>0.006439</td>
<td>2.61*</td>
</tr>
<tr>
<td>Hackamount</td>
<td>2.72e-07</td>
<td>-0.25</td>
</tr>
<tr>
<td>Herfindahl</td>
<td>3.64e-06</td>
<td>0.27</td>
</tr>
<tr>
<td>Wald chi2(19)</td>
<td>902.95*</td>
<td></td>
</tr>
</tbody>
</table>

Based on the result, it is obvious that only the Hacking Occurrence (dummy variable) demonstrated statistical significance at 95% confidence. This signals that the lagging effect to the
Bitcoin market volatility is strongly and linearly correlated with the occurrence of the hacking event. Whenever hacking activity shows in the public media, the market will react and consequently digest the shock in the list of moving average days following the publication date. Furthermore, the amount of hacking does not correlate with the length of volatility. The market does not produce longer volatility about the quantity of Bitcoin involved in the hacking activities. However, because there does not exist an expression for expectations of hacking, such linear correlation is not shown with ARMA function. However, as most of the events were collected in the first two years, this would require a separate study to focus on hacking.

Fitting the expected value to the original Price change line graph, it is visible how ARMA can model the volatility of the Bitcoin price: although only by a small percent, ARMA’s fitted value fits within the original volatility diagram and they behave in the same fashion.

The volatile parts of the ARMA prediction correspond to the actual Bitcoin price volatility. The red section in this combined line graph is the part of Bitcoin volatility that is caused entirely by its past and the volatility that is caused by the lagged shock from each of the hacking news. Running an OLS with two variables, I have found a t-value of 10.69 and adjusted $R^2$ of 0.05. This further demonstrates that ARMA-generated data significantly forecasts the volatility of Bitcoin, although the ARMA effect only accounts for approximately 5% of total volatility in the Bitcoin market.
The residuals from the ARMA model are mostly free from autoregressive effect. Those residuals are slightly different from the original price volatility data. Before applying the GARCH model, I have run Engle’s Lagrange Multiplier to ensure that ARCH effect still exists. The result shows that ARCH disturbance still exists for lags up to 30. It is sufficient to apply GARCH (1, 1) as the minimum parameter. Because the residuals are from the ARMA model, they are free from moving average lag; thus, they will most likely not display higher order GARCH effect. Furthermore, Chu et al. (2014) have shown success with GARCH (1, 1) modeling. For these reasons, I will directly apply GARCH (1, 1) for the analysis. The result is as follows:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Z score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hackoccure</td>
<td>-0.0344012</td>
<td>-16.92*</td>
</tr>
<tr>
<td>Hackamount</td>
<td>4.23e-07</td>
<td>8.61*</td>
</tr>
<tr>
<td>Herfindahl</td>
<td>-3.46e-06</td>
<td>-2.52*</td>
</tr>
<tr>
<td>Wald chi2(3)</td>
<td>310.72*</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: GARCH Results

The high Wald $\chi^2$ result shows that, although free from autoregressive disturbance, the GARCH (1, 1) model still fits well and yields significant results. All coefficients are significant at 95% confidence level. This indicates that both hacking, and mining concentration affected the price volatility.

A higher Herfindahl index value, which indicates a lower miner diversity, or a higher firm concentration, strongly reduces the price volatility. Surprisingly, previous literature has illustrated the increased risk with a more concentrated miners’ distribution and the potential of collusion that would destroy the Bitcoin system. This negative relationship between Herfindahl and volatility calls for reinvestigation of the game theory between the miners. One possible reason for this phenomenon is that, as the mining difficulty climbs over the years, the cost of
mining includes more than electricity. The portion of the fixed cost, such as investments on the mining card and real estate may have increased to the degree that altered the disposition of the miners, making exiting market incredibly costly. This low potential of colluding could be the reason for such negative relationship. Another possible reason is that, as the top companies become predominant, they have a stronger voice to stabilize the price change because of their supply power.

Hacking activities also demonstrate a strong and significant influence on the Bitcoin price volatility. The positive coefficient of hacking amount shows that, as more Bitcoins are hacked, scammed or stolen, volatility in the Bitcoin price increases following the event. However, the negative coefficient on hacking occurrence indicates that the hacking amount is not linearly correlated to the magnitude of volatility. Because of the positive correlation between hacking occurrence and amount, if Bitcoin price volatility linearly increases as hacking amount increase, the coefficient of hacking occurrence should be positive and small. This non-linear relationship cannot be explained using GARCH modeling. Since hacking occurrence and hacking amount are values at the event dates and all other entries are zero, hacking occurrence thus serves as a weight to the variable.

Regressing the residuals from the GARCH process with the original price change percent, OLS yields an $R^2$ value of 0.9496. This shows that the Hybrid ARMA-GARCH model with 3 variables was able to predict 5.04% of the total volatility of Bitcoin price in the past 6 years.
Conclusions and Implications

With the high Wald $\chi^2$ value, this model has successfully yielded three major findings. This section will present those findings and discuss potentials for future research. The three findings are as follows:

**Non-Seasonality effect on Bitcoin price volatility**

The autoregressive effect of Bitcoin price volatility is significant. Based on the ACF and PACF graphs, volatilities last up to 30 days on the Bitcoin market. However, such effect is not consistent throughout the 5-year observation period because the outliers in the graph do not converge in a linear fashion.

Furthermore, it was speculated that Bitcoin would have a seasonality effect at a seven-day lag because Bitcoin is traded with multiple currencies every day including weekends, whereas currencies and stock markets only operate on weekdays. Based on the result from the Autoregressive model and AIC test, such discrepancy does not cause volatility to have a weekly effect. Because of this, the Bitcoin market will hardly have “Black Friday” or “Black Monday” effects that is commonly observed in other financial markets.

**The relationship between hacking and Bitcoin price volatility**

The result from the ARMA process shows us that hacking events have a significant effect to the lag of Bitcoin price volatilities. That is, each hacking event makes the Bitcoin price more volatile for a few days after. In particular, Bitcoin price volatility responds most significantly at day 1, day 2, day 3, day 5, day 6, day 10, day 15, and day 24 following the occurrence of the event. Interestingly, volatility lags respond only to the occurrence of each hacking event but not
to the amount involved. The amount of Bitcoin involved in each hacking event does not have any linear relationship to the magnitude of Bitcoin price volatility.

The GARCH process shows us that both the occurrence and the amount of Bitcoin involved in each Bitcoin hacking event have significant effects on the Bitcoin price volatility. Because of the high standard deviation in the hacking amount variable, the occurrence of the Bitcoin hacking is more significant than the amount involved in the price volatility.

**Miner centralization to Bitcoin price volatility**

The ARMA-GARCH model shows that miner centralization does have a statistically significant effect to the volatility of Bitcoin. From the Z score difference, the effect of centralization change is less significant than the effect of hacking events. As speculated, a higher Herfindahl Index value, or a less diversified miner market, results in a lower volatility of Bitcoin price. However, because of the small coefficient of Herfindahl, market centralization’s effect on Bitcoin price volatility is minimal.

This study has provided an exploratory study on the effect of Hacking and miner centralization to Bitcoin price volatility. Although the independent variables are statistically significant, and the model generates a high Wald $\chi^2$ value, the forecasting power for the actual Bitcoin price volatility is limited. This study can be strengthened in many ways.

The data of Bitcoin hacking events are manually collected and applied to the model directly. Such a data collection method is insufficient for the model. Future research may seek alternative methods of building the history of Bitcoin hacking.
A USD value for Bitcoin stolen may be more accurate to match with market speculators’ expectations. As the time-sensitive sample grows over time, an inflation-adjusted measurement would also be more appropriate in future studies.

This paper used BTC.com as the source for miners’ distribution data, which may not be the ideal way to measure centralization. Although the assumption that Bitcoin miners are the owners of each mined block is sound, the assumption that unknown miners are equally distributed is weak. A better method of measuring the ownership of the blocks could be to parse the blockchain. By sorting the IP addresses of each mined block, the measurement of miners’ distributions can be much more accurate.

The hybrid ARMA-GARCH model is successful in accommodating all variables. Because of the limited sample collection, the forecasting power of the model is limited, as shown in the small coefficient. The model has used all three variables in both stages. In future research, with a better understanding of the effect of the independent variables, the ARMA process can adopt the variables that are predicted to cause a lagging effect to the dependent variable. Removing the irrelevant terms from the detrending process can increase the accuracy of model’s forecast.

This paper offers new views on the Bitcoin price volatility change. For Bitcoin investors, a pattern of frequent hack, scam, and loss news is a strong signal of an increased volatility in the following three weeks. The evidence gathered from this paper strongly suggests that the Bitcoin trader must observe market fluctuation at all time, including weekends. Although both data and literature suggest that Bitcoin price volatility has been decreasing over time, some inherent properties including the miners’ distribution suggest a trend of increase. When speculative bubble decreases in the future, the stochastic models may yield a stronger forecasting power.
Work Cited


Retrieved from: https://blockchain.ino/pools?timespan=24hours


Cheng, E. (2017) Dark web finds bitcoin increasingly more of a problem than a help, tries other digital currencies. CNBC.


Retrieved from: https://trends.google.com/trends/explore?date=all&q=bitcoin


Work Consulted


