THESIS

BITCOIN PRICE FORMATION: AN EMPIRICAL INVESTIGATION

Submitted By

Aric Light

Department of Economics

In partial fulfillment of the requirements

For the Degree of Master of Arts

Colorado State University

Fort Collins, Colorado

Summer 2019

Master's Committee:

Advisor: Daniele Tavani

Anita Pena
Stephan Kroll

ABSTRACT

BITCOIN PRICE FORMATION: AN EMPIRICAL INVESTIGATION

Created in 2008 and rising to prominence in 2017, Bitcoin continues to generate controversy as to whether it is a speculative asset or the harbinger of a future of global, decentralized commerce. The focus of this paper is to investigate the properties of Bitcoin and its market by assessing asset specific factors (users, hash rate, etc.) and traditional market factors (market risk, currency risk, etc.). The objective is to quantify the impacts of these forces as drivers of Bitcoin returns and to develop a risk measurement framework with the potential to inform future use cases.

The analysis is broken out into two parts. The first seeks to quantify the impact of asset specific "supply and demand" factors with respect to Bitcoin's daily price return and volatility and to determine the relative efficiency of the nascent Bitcoin market. To do this a GARCH model is specified which enables the measurement of return impacts and volatility within a single model. A back test and forecast are then conducted to determine if the conditional value at risk and expected shortfall can be accurately captured by the model. We determine the Bitcoin market is weakly efficient, returns are highly impacted by supply and demand factors and that the specified value at risk model accurately describes the exceptional volatility.

The second part incorporates a set of macro-financial variables into the model to determine Bitcoin's exposure to traditional sources of risk; such as stock and currency market returns. The results show that Bitcoin is largely unimpacted by broad macro-financial variables once supply and demand variables are properly accounted for. This suggests that although Bitcoin is a weakly efficient market it is generally disconnected from worldwide capital and currencies market. This further suggests that Bitcoin may have currently limited "real world" use cases which is an important consideration for investors.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Initially developed in 2008 by Satoshi Nakamoto and described in what has become his/her seminal paper *Bitcoin: A Peer-to-Peer Electronic Cash System*[1], Bitcoin has become both an object of controversy and fascination in recent years and was the catalyst in what has become the burgeoning space of cryptocurrency. Bitcoin is an unregulated, digital asset that is not issued as physical currency and, unlike world currencies, is not associated with a government nor is it issued by a central monetary authority. The system instead relies on what is dubbed a "decentralized distributed ledger" that is maintained by blockchain technology. The blockchain is unique in that it is uses a cryptographic proof of work algorithm to verify transactions on the Bitcoin network thus enabling users to send payments near instantaneously while ensuring that payments are legitimate and subsequently preventing the possibility of "double spending".

Since its inception late in 2009, the USD exchange rate for Bitcoin (BTC) peaked on December 17th, 2017 at $19,982.60/BTC. As of this writing there are an estimated 1,658 registered cryptocurrencies trading on 200 exchanges worldwide with a combined market cap of approximately $128B. Bitcoin continues to dominate the space with a market cap of $68B; representing 53.1% of the total market value[2]. The space has experienced a remarkable degree of maturation in a short period of time but is still quite small in comparison to traditional equity and debt markets which, as of the end of 20l7, were estimated to have global market capitalization rates of $70T and $92.2T, respectively.[3]

While cryptocurrency and Bitcoin have been the subject of substantial speculation, it has yet to make any identifiable headway into the banking system or everyday commerce. It is estimated that in the fourth quarter of 2017 only 11,291 businesses worldwide accepted Bitcoin as a valid form of payment; the substantial underground Bitcoin economy notwithstanding.[4] Additionally, banks have been reticent to adopt the use of cryptocurrency primarily due to the ambiguous regulatory framework and lack of traditional KYC (Know Your Client) guidelines. This is further complicated by confusion and speculation as to the nature and use of Bitcoin. Given that so few businesses currently accept it as payment it can hardly be said to be functioning as a currency in the traditional sense leading some to classify it simply as a speculative or (at best) a hedging asset like gold or other precious Earth minerals.[5]

For these reasons there has been a lack of adoption within the sub-industry that might seem the most natural fit for Bitcoin: financial services. The objective of this paper is to fill in those gaps and to provide empirical insight into the nature of Bitcoin as a financial asset. Thus, the objectives of this paper will be 2-fold: 1) to provide an assessment of the market efficiency and financial time series characteristics of Bitcoin and accurately model the characteristic volatility using an ARCH/GARCH model, 2) to assess Bitcoin's exposure to traditional macro-financial and asset specific variables which could plausibly be driving the return generating process. By accurately describing Bitcoin's statistical properties and characteristics it becomes possible for individuals and financial institutions to develop appropriate use cases such as: allowing Bitcoin to serve as eligible collateral or issue Bitcoin denominated debt.

The paper will be organized as follows: chapter 2 will discuss the relevant literature, most of which is quite recent given Bitcoin's relative novelty, chapter 3 presents an overview of how the Bitcoin/blockchain technology works in order to provide readers with context for the analysis to follow, chapter 4 reviews the data sources and software used, chapter 5 concerns the first objective of study and evaluates Bitcoin's daily price return and volatility series using variables unique to the network and discusses the results in the context of efficient market theory, while chapter 6 analyzes the macro-financial exposure of Bitcoin using weekly time series data.

CHAPTER 2: LITERATURE REVIEW

Bitcoin's ascent has generated a growing interest in its academic study. Yermack (2015) was among the first to assess whether Bitcoin qualifies as a currency; namely, does it serve as 1) a medium of exchange, 2) a store of value, and 3) a unit of account.[5] He concludes that it serves weakly as a medium of exchange citing that it can be used to purchase goods, but has practical limitations involving a lengthy procurement process and "spending time"; since transactions need to be verified algorithmically, using BTC to pay for a product takes significantly longer than a traditional paper currency or credit card. It serves as a very poor unit of account due to its extreme volatility which would compel retailers to adjust prices frequently. Additionally, markets are not uniform and different exchanges often quote BTC at different prices. Finally, it also serves poorly as a store of value. Due in part to the extreme volatility, BTC does not make a desirable safe harbor for storing money long term and furthermore is not accepted broadly as eligible collateral. Yermack also cites the prevalence of cyber-security threats and their unconformably high degree of success against BTC's exchange network. Additionally, the storage systems for BTC are archaic and involve multiple layers of security.

Glaser et al. (2014) examines the differences between intra-network and on exchange volume to determine if Bitcoin should be classified as a currency or financial asset.[6] By examining transaction and exchange volume since Bitcoin's inception, the authors posit that increased adoption and awareness of Bitcoin should translate to both increased exchange and transaction volume if BTC primarily functions as a currency and only on-exchange volume if it behaves more like a financial asset. Using Wikipedia term search and Google Trends data the authors attempt to determine how interest in BTC relates to exchange and transaction volume respectively. They conclude that interest in BTC has led to an increase in exchange trading volume but has not had a statistically significant impact on network transaction volume which suggests that BTC functions less like a currency and more like a speculative asset.

Hayes (2017) takes an engineering-based approach to constructing a valuation model for Bitcoin and 65 of the other most popular cryptocurrencies.[7] The author posits that cryptocurrency value can be derived based on the difficulty of mining a new block (thereby creating the coins) proxied by the computational power available, the computational algorithm used, the number of coins found per minute, the percentage of coins mined thus far, and life of the network. Using a simple regression model, the author finds that computational difficulty, coins per minute and the algorithm used have a statistically significant impact on the price of the cryptocurrencies used in the cross section and the regression has a high degree of explanatory power.

Jeffrey Chu, Saralees Nadarajah, and Stephen Chan (2015) present a very comprehensive analysis of the daily returns for the BTC-USD exchange rate.[8] The authors claim that when dealing with daily return data the normality assumption for statistical inference is often violated; as documented for stock returns by Stephen J. Brown and Jerold B. Warner.[9] The authors proceed to fit 15 of the most widely applied parametric statistical distributions to daily price returns for the BTC-USD exchange rate and conclude that a generalized hyperbolic distribution best fits the data. This distribution suggests that the Bitcoin returns exhibit substantially heavier tails and more peakedness than would be expected from a normal distribution. This selection was empirically determined based on information criteria which penalize for the inclusion of additional parameters.

Urquhart (2016) employs the familiar Ljung-Box and runs test to assess if Bitcoin returns follow a simple random walk process.[10] Using daily data spanning January 2010 to July 2016 the author concludes that Bitcoin returns demonstrate significant inefficiency over the full sample period. He proceeds to divide the sample into two equal sized halves spanning January 2010 to March 2013 and April 2013 to June 2016 respectively and re-runs the tests. The test results show that Bitcoin returns demonstrate inefficiency in the first sub-sample but appear weakly efficient in the latter sample period. He concludes that for a majority of its life through 2016 Bitcoin was an inefficient market but is in the process of becoming an efficient market due to increased user adoption.

As a follow up to the Urquhart study, Chu and Nadarajah (2017) examine an expanded set of tests to evaluate the relative efficiency of Bitcoin returns over the sample period January 2010 to July 2016, but first clean the data to remove outlier points.[11] Their test results show that Bitcoin returns are weak form efficiency over the full sample period if such points are removed.

Feng et al. (2018) also investigate the question of efficiency by testing for the presence of informed trades in order level data.[12] They define informed trades as large blocks of trades placed prior to market events that should not be known in advance (hacks, regulatory announcements, etc.). They estimate the effect using the imbalance between buy and sell orders observed on the exchange prior to a given event. The rationale follows that if informed trading is taking place, then we should observe a negative (positive) order imbalance prior to the release of negative (positive) information; i.e. more traders trying to sell (buy) ahead of negative (positive) news. They conclude that significant order imbalances in both directions are observed prior to 42 cited events and that substantial informed trading is taking place. The authors do not specifically opine on the implications of their conclusions for market efficiency.

Bariviera, Basgall, Hasperuéb, and Naiouf (2017) investigate the presence of what they call "long-memory" (i.e. non-zero correlation going many periods back in time) in the sequence of daily returns for the BTC-USD exchange rate beginning at inception in 2011 through 2015.[13] The detection of "long-memory" would suggest a violation of the Efficient Market Hypothesis and signal inefficiencies in the incorporation of information about BTC into price. Using Hurst exponent analysis, the authors show the existence of "long-memory" in the BTC-USD exchange rate from its inception in 2011 through 2014. However, around 2014 the dynamics appear to change, and the exchange rate behaves like a random walk; furthermore, the Hurst exponent converges to .5 which is consistent with such a finding. They compare this behavior to other major currencies including the Euro and British Pound (each have a calculated Hurst exponent of .5) to show the similarity in behavior.

Jana et al. (2018) document similar findings with respect to the information efficiency of Bitcoin.[14] They extend the work of Bariviera et al. by employing a battery of long-range dependence estimators for a period spanning July 18, 2010 to June 16, 2017. Their findings suggest that the Bitcoin market is general efficient, but has exhibited several, relatively short periods of inefficiency; specifically, the periods spanning April to August 2013 and August to November 2016.

Van Wijk (2013) attempts to evaluate Bitcoin's exposure to macro-financial variables in both the short and long-run.[15] The author specifies an ECM model and includes a cornucopia of familiar macro-economic indicators including the Dow Jones Industrial Average, the Nikkei 225, FTSE 100, Brent oil price, WTI oil price, and several world exchange rates among other variables. Based on t-tests, the author concludes that the Dow Jones, WTI oil and USD-EUR exchange rate significantly impact the price of Bitcoin in the long-run while only the Dow is significant in the short run.

Kristoufek (2013) documents a link between investor interest and Bitcoin price.[16] He specifies 2 different forecasting volumes to gauge the impact of search volume from 2 different sources. Model 1 specifies a simple VAR (1) and focuses on the price response based on Google search data (as prioxied by a Google Trends index where 100 denotes peak interest). Model 2 specifies a VECM and utilizes Wikipedia search volume measured using daily searches. The Wikipedia results are particularly interesting as they demonstrate a significant positive feedback between the price innovation of BTC and search volume: If the prices are going up and interest is growing, then the price reacts positively. But if the price declines, then the increased interest pushes them even lower.

Ciaian (2016) proposes a series of structural VAR models for Bitcoin based on an augmented version of Barro (1979) .[17,18] Using both level and price return data for BTC from November 2009 to May 2015 the authors assess 3 different hypotheses: 1) the Bitcoin supply and demand drivers, 2) Bitcoin's attractiveness as an investment, and 3) its exposure to macro financial variables. Bitcoin supply is unique in that it is established by the source code and is predetermined and hence can be treated as an exogenous variable. The authors investigate hypothesis 1 by examining demand side drivers of Bitcoin price including the size of the BTC economy as proxied by the number of unique addresses on the network, number of transactions, and the velocity with which BTC changes hands. They find the impacts of the number of transactions and network addresses to be positive and statistically significant while, contrary to expectations and theory, find the velocity to have a negative impact and be insignificant.

Bitcoin's investment attractiveness is evaluated based on the volume of Wikipedia search queries related specifically to BTC as well as the number of new members and new posts extracted from the popular crypto forum, bitcointalk.org. They conclude that, in the short term, Wikipedia search data as well as new members and new posts have a statistically significant impact on BTC price, but that only the variable *new posts* is significant in the long run. The authors suggest that in the long run information on BTC has more time to circulate and thus are not surprised that the effects of Wikipedia views and new membership diminish in the long run. They also offer the possible caveat that these variables may be indicative of more than just the interest of potential investors and likely also reflects increased interest in BTC education more generally which would not impact price and may be distorting the magnitude of the effect.

Finally, using the value of the Dow Jones Industrial Average, price of Brent crude oil, and USD-EUR exchange rate, the author's attempt to capture BTC's exposure to broad macro-financial conditions. They conclude that these typical macro-financial variables do not impact the price of BTC in the long-run as none indicate any degree of statistical significance at conventional levels. These results contradict the earlier results of Van Wijk which lead the authors to speculate that the results of Van Wijk are biased and that supply-demand effects feature more prominently in the determination of Bitcoin price.

Dyhrberg (2016) studies the Bitcoin return time series in a GARCH framework with the goal of drawing conclusions about the currency and (once again) its exposure to macro-financial variables.[19] She incorporates the returns of the FTSE 100, gold futures, USD-EUR and USD-GBP exchange rates, and the Fed Funds rate as independent variables. She proposes two, time series, GARCH models for describing the behavior of the BTC price return series. The first is the standard GARCH (1,1) model and the second is the Exponential-GARCH (E-GARCH) (1,1) model initially devised by Nelson (1991)[20] that accounts for potential asymmetry in the volatility equation. That is, volatility may adjust asymmetrically in response to positive or negative news; this is also known as the leverage effect and is a well-documented phenomenon found in many financial time series. She advocates including all the variables in both the mean and volatility equations. Both models assume a standard normal distribution of the residuals.

Results from the models indicate the BTC responds significantly to the Fed Funds rate, changes in the USD-EUR and USD-GBP exchange rates, and FTSE 100 in both the mean and volatility equations. Interestingly, the results suggest that BTC volatility decreases in response to a positive shock in the Fed Fund rates; an unusual response considering the dependency documented between asset prices and prevailing interest rates. Moreover, BTC volatility is also shown to decrease following a positive change in the USD-GBP exchange rate which the author suggests may signal possible hedging and risk management opportunities. Furthermore, she documents the presence of leverage as Bitcoin volatility increases more in response to "bad news" (i.e. increases following a negative surprise or innovation).

Bouri (2016) the authors propose a dynamic correlation model based on Engle (2000) to evaluate the usefulness of Bitcoin as a portfolio diversifier, hedge, or safe-haven asset.[21,22] The author defines a diversifier as an asset that is either uncorrelated or weakly positively correlated with another asset, a hedge as an asset that is uncorrelated or negatively correlated, and a safe-haven as uncorrelated or negatively correlated with other assets specifically in times of "stress". Gold has often been characterized as having all three of these qualities thus making it attractive in a variety of economic environments. The authors assess these three capabilities against many capital market indices (S&P 500, FTSE 100, Nikkei 225, MSCI World, etc.), the US-Dollar exchange rate index, gold, and Brent crude oil futures using daily and weekly data from July 2011-December 2015. Daily results suggest that BTC is not an effective safe-haven during turbulent periods for any asset, but that for many assets (including the Nikkei 225 and commodities) it is a reasonably good hedge. Moreover, for all assets studied, Bitcoin serves as an effective diversifier due to its low unconditional and conditional correlation with these assets. However, these possible hedging and diversification benefits are eroded to a considerable degree when calculated for weekly data presumably because Bitcoin features continuous trading and most other financial assets only trade Monday-Friday.

In possibly the most comprehensive paper to date on the relative efficiency of the Bitcoin market, Bitwise Asset Management analysts Matthew Hougan, Hong Kim, and Micah Lerner examine the impact fake volume and fake exchanges in the BTC market.[23] In conjunction with Google, the authors build a custom data set that aggregates reported trade data from 83 of the top BTC exchanges worldwide. Based on a highly granular analysis of trade size, trading patterns and spreads and authors conclude that 73 of the 83 exchanges and an astonishing 95% of reported data is fake. Furthermore, they suggest that among the 10 legitimate exchanges a highly efficient market is observed. Their conclusion is based on an examination of trading spreads and the timing of trades. The 10 legitimate exchanges show a high degree of overlap between when trading takes place and the size of trades which implies traders across platforms react similarly in response to market developments. Additionally, they document consistent and tight spreads across platforms which implies a functional and highly liquid market. They conclude that although substantial false information and market action exists within the BTC market, the effect on price is minimal and any deviations across exchanges is quickly arbitraged away.

CHAPTER 3: STYLIZED FACTS ABOUT BITCOIN

In this section we will briefly discuss some of the characteristics of Bitcoin that make it unique as a currency. Bitcoin supply is exogenously determined as the number of new coins and the time interval at which they are brought to market is set in advance by the algorithm that governs the network; a process commonly referred to as mining. Miners are essential to the Bitcoin network as they are responsible for verifying the validity of transactions and recording ownership of the available Bitcoins at a given point in time. Miners verify transactions by solving a mathematical algorithm; in Bitcoin's case the algorithm is Standard Hash Algorithm (SHA)-256. The answer consists of a time stamp indicating when the transaction took place and the public key of the previous owner; this process is typically referred to as "hashing" and essentially involves a slightly more elegant version of guess-and-check. The more powerful the computer that is used in the hashing process, the more guesses it can make per second; this is known as the "hash rate" and is typically measured at the network level. Once a transaction has been verified it is added to a "block" where the record is permanently encoded into the network and made publicly available. Historically an average of 2020 transactions have been recorded on a single block, but this will vary over time depending on the size and number of transactions taking place on the network as well as the hash rate.

Mining is a very computationally difficult and energy intensive process. To incentivize miners to join the network and mine new blocks a reward mechanism was devised that serves the dual purpose of incentivizing participation and increasing the Bitcoin money supply. Once a new block is found the miners who participated in the formation of the block are rewarded a specific number of Bitcoin as payment which is divided proportionately. As of this writing the current reward for mining a new block is 12.5 BTC or approximately $50,000; a value which fluctuates depending on the current exchange rate. The block reward is not fixed and is halved every 210,000 blocks as the total Bitcoin money supply approaches the maximum of 21MM BTC in circulation. The governing algorithm is constructed in such a way that a new block is found every 10 minutes (i.e. block time). In this way, the Bitcoin network can bring a fixed amount of new money into circulation at fixed intervals.

CHAPTER 4: DATA AND SOFTWARE

This study will be concerned with addressing two objectives related to the financial time series properties of Bitcoin: 1) an assessment and characterization of Bitcoin's pure daily return and volatility, and 2) Bitcoin's exposure to macro-financial and currency specific properties. Bitcoin has unique features as a financial asset in that all market transactions are executed digitally and there is no physical exchange (though that is in the works). A consequence of this is that Bitcoin is traded around the clock and exchanges are never closed. This contrasts with other financial assets that trade (at least partially) on physical exchanges and exclusively on a traditional Monday-Friday schedule, between specific times and are often closed on holidays.

All data for the Bitcoin price series was obtained from the Bitstamp exchange. Bitstamp was founded in August of 2011, is headquartered in Luxembourg and was selected based on its popularity among European and US customers and its time in existence. Furthermore, the Bitwise study cited Bitstamp as 1 of the 10 legitimate exchanges it surveyed. The time frame under consideration varies slightly depending on the objective.

- Objective 1 was conducted using daily data from September 14, 2011 through March 28, 2018 for BTC and the network specific variables; all of which were available at the daily frequency.
- Objectives 2 incorporates additional variables including stock market indices, exchange rates and interest rates which either do not trade on weekends or data is simply not available for those times. As such, this objective relies on weekly data. BTC weekly returns are generally not available and it was critical that the time step for both BTC and other model variables be as similar as possible. The BTC weekly return data and network variables data were therefore calculated from Saturday-Friday to incorporate the full return series and match the limitations of the other variables as closely as possible. Log returns for the other model variables were calculated as Monday-Friday or, in the case of interest rates, a level reading was taken on Friday each week.

Data for exchange rates were obtained from Quandl; a popular, free online database used in several of the cited studies.[24] Market data was accessed through Yahoo! Finance and has been adjusted for dividends and splits.[25] Interest rate figures were gathered from the Federal Reserve Bank of St. Louis' FRED database.[26] The time frame under consideration for these objectives spans from January 6, 2012 through March 23, 2018.

All computations presented in this study were performed in **R** and make extensive use of the packages available therein or through the CRAN network. Most notably used was *rugarch* developed and maintained by Alexios Ghalanos.[27]

CHAPTER 5: OBJECTIVE 1-ANALYSIS OF DAILY BITCOIN RETURNS AND VOLATILITY

Section A: Econometric Approach

Objective 1 concerns the estimation and evaluation of Bitcoin's daily return and volatility using a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. Autoregressive Conditional Heteroskedasticity (ARCH) models were first proposed by Engle (1982) in which the current volatility of a stationary process is modeled as a function of its past innovations.[28] The ARCH model was later generalized to GARCH by Bollerslev (1986) who proposed modeling volatility as a function of past innovations and the past realized volatility of the process.[29] The general form of GARCH can be written as:

$$\Delta Y_t = \theta_0 + \sum_{i=1}^{p} \theta_{t-i} Y_{t-i} + \sum_{j=1}^{q} \phi_{t-j} \varepsilon_{t-j} + a_t$$

where

$$a_t = \varepsilon_t \sigma_t$$

$$\sigma_t = \sqrt{\alpha_t + \sum_{m=1}^{M} \alpha_m a_{t-m}^2 + \sum_{n=1}^{n} \beta_n \sigma_{t-n}^2}$$

where

$$\varepsilon_t \sim N(0,1)$$

ARCH and GARCH models have become some of the most dominant and widely applied models employed in statistical finance and have spawned a great number of models with heightened complexity.

GARCH models are applicable to stationary time series. Figures 1 and 2 depict the Bitcoin-USD exchange rate and the daily return time series respectively.

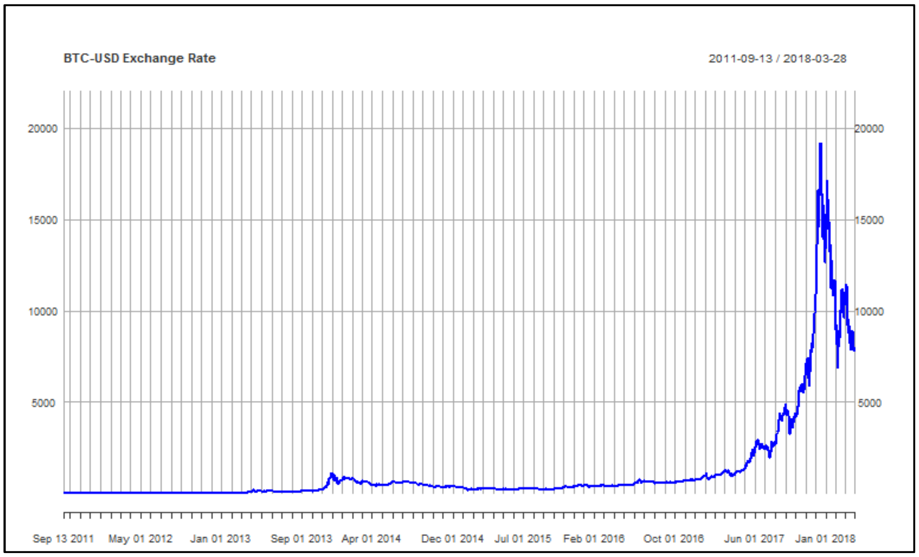

Figure 1: Daily closing price series for Bitcoin Jan. 6, 2012 to March 23, 2018.

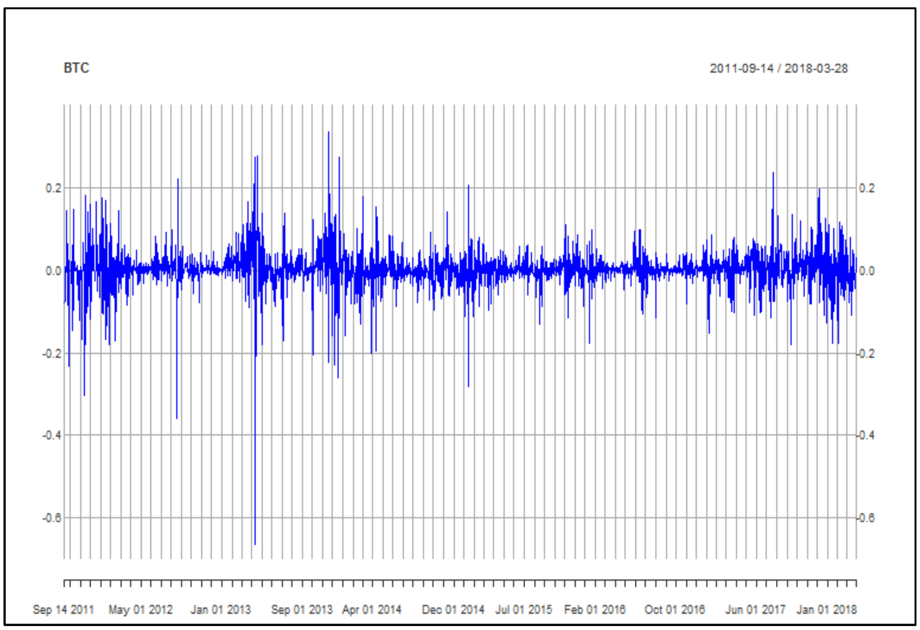

Figure 2: Daily adjusted return series for Bitcoin Jan. 6, 2012 to March 23, 2018.

Visual inspection of these graphs strongly suggest that the Bitcoin-USD exchange rate is non-stationary in levels but is stationary in first differences. To confirm this intuition, we employ the Augmented Dickey Fuller (ADF) and Phillips-Perron (PP) tests to examine for the existence of unit roots in the time series'; the results for which are presented in Table 1 along with some descriptive statistics for the time series'.[30,31]

Table 1: Descriptive Statistics and Unit Root Tests

Statistic	BTC-USD Exchange Rate	BTC Price Return
Observations	2388	2388
Mean	1231.57	.002
Median	374.68	.00199
Minimum	0	-.664
Maximum	19187.78	.3374
Std. Deviation	2776.31	.0503
Skewness	3.5938	-1.311
Kurtosis	13.425	20.389

Statistic	BTC-USD Exchange Rate	BTC Price Return
Augmented-DF τ-statistic	-1.9503	-11.45
ADF p-value	.5994	<.001
Phillips-Perron τ-statistic	-7.9021	-2449.7
PP p-value	.669	<.001

The ADF test statistic follows the tau (τ) distribution developed specifically for this test. The PP test uses non-parametric methods, but the test-statistic has been shown to be asymptotically distributed (τ) which we appeal to given the relatively large sample size of 2388. Both tests have a null hypothesis of non-stationarity. Results from the tests confirm the existence of a unit root in the BTC-USD exchange rate time series but suggest that the price return series is stationary.

Visual inspection of the price return plot also appears to suggest the presence of volatility clustering. First noted by Mandelbrot (1963)[32], volatility clustering refers to the time series phenomenon whereby "…large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes." Squared residuals around the mean can be used to proxy the (possibly heteroskedastic) volatility of the process and assess the existence of clustering; Figure 3 presents this for the BTC price return series.

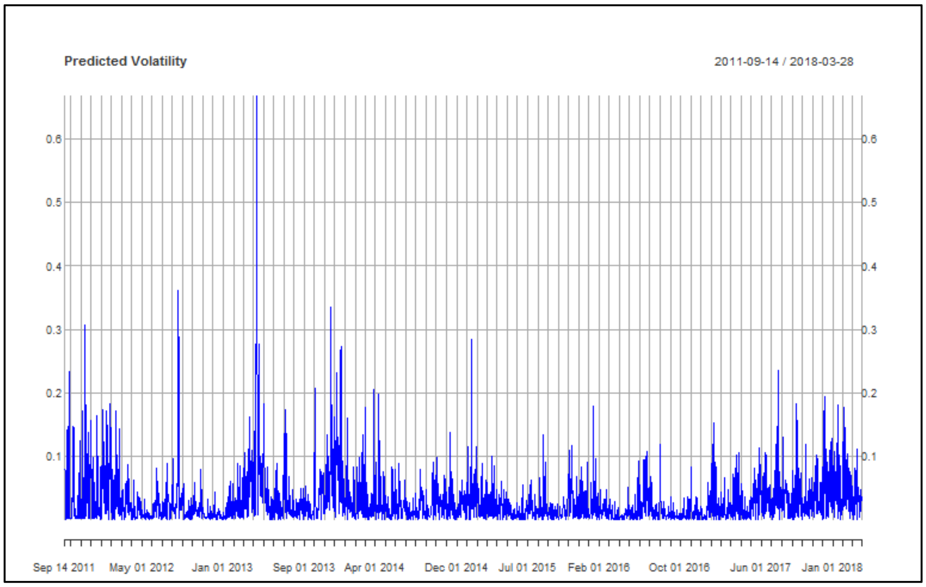

Figure 3. Time series of square root of squared residuals

We employ the ARCH LM Test introduced in Engle (1982) to formally test for the presence of ARCH effects in the time series. This joint test evaluates for ARCH effects up through lag order p. The LM test statistic is asymptotically distributed χ_p^2 under the null hypothesis of no ARCH. The test statistic can be derived by regressing the model residuals, $\hat{\varepsilon}_t^2$, on residuals up through lag length p, $\left(\hat{\varepsilon}_{t-1}^2 \dots \hat{\varepsilon}_{t-p}^2\right)$. The test statistic is derived as follows where R^2 is the coefficient of determination from the auxiliary regression.

$$n * R_{asy}^2 \sim \chi_p^2$$

Table 2 presents the statistic and associated p-value for the test for selected lag lengths:

Table 2: ARCH test statistics and p-values for BTC price return

Order(p)	LM Test Statistic	p-value
4	2511	<.001
8	1127	<.001
12	746	<.001
16	550	<.001
20	436	<.001
24	354	<.001

Results indicate significant ARCH effects that decay very slowly. Such results suggest that an ARCH or GARCH model should be employed to model this time series.

We propose a GARCH model for modeling the BTC price return time series that takes advantage of the unique supply and demand characteristics of Bitcoin previously cited in the literature. As previously mentioned, the growth of the monetary base of the Bitcoin network is exogenously determined. However, it may still make sense to include "supply" variables in our model to account for the needs of the network. Recall that the hash rate (measured in giga-hashes per second (GH/S) i.e. billions of guesses per second) represents the total capacity of the network to verify transactions and mine new blocks. This is a critical measure of the BTC ecosystem as it represents the amount of computing power available to keep the network functional given the number of transactions and exchanges taking place. Consequently, even though the block time and block reward are fixed, if the hash rate is low relative to the network demands then very few transactions will be verified and included in each block. This would induce illiquidity and inefficiency in the BTC network and potentially lead the system to crash. Thus, it seems prudent to include the hash rate in our model to proxy for network capabilities and sustainability.

The literature has shown that demand side factors are an important component of the BTC price determination. Kristoufek (2013) and Pavel Ciaian, Miroslava Rajcaniova & d'Artis Kancs (2015) advocate the use of Wikipedia and Google search volume as a means of measuring Bitcoin's attractiveness as an investment vehicle.[16,17] Glaser et al. (2014) suggests using exchange volume and transaction volume to isolate the effects of real spending and investment on price determination.[5] Both are interesting and should be considered in conjunction. However, some difficulties were encountered when trying to attain search related data. Specifically, Wikipedia recently revamped their API interface and as such data on traffic to the Bitcoin or related pages in only available going back as far as 2015 and the old archives appear to be either removed or rendered inaccessible. Furthermore, Google does not track daily search volume for BTC and only offers weekly data going back to 2013 and monthly data thereafter. The Google search data has also been manipulated in such a way that current and past search interest is reflected as a percentage of peak interest (which appears to have occurred in December 2017). Additionally, these are dirty measures rife with issues regarding the extent to which searches translated to increased adoption versus general education or simply passing interest. These issues represent significant limitations on the incorporation of "interest" variables into the model and as such will not be discussed here.

Network and exchange data however are widely available and spans the period under consideration. Following the advice of Glaser, we have elected to include both exchange and transaction volume for Bitcoin to capture the impacts of investment only versus real price effects. Additionally, we have elected to include the number of unique addresses (i.e. users) registered on the Bitcoin network to assess the growth in adoption of Bitcoin more broadly.

The final model can be specified as follows:

$$r_{BTC,t} = \beta_0 + \theta_1 r_{BTC,t-1} + \theta_2 r_{BTC,t-2} + \beta_1 ExVol_t + \beta_2 TrVol_t + \beta_3 Ads_t + \beta_4 HR_t + a_t$$

where

$$a_t = \varepsilon_t \sigma_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_{t-1}\sigma_{t-1}^2 + \lambda_1 ExVol_t + \lambda_2 TrVol_t$$

and

$$\varepsilon_t \sim GED(0,1,\kappa)$$

Where:

- $r_{BTC,t}$ is the price return of Bitcoin at time t. And the associated lags at *t-1* and *t-2*.
- $ExVol$ is the daily change in the volume of transactions occurring on exchanges.
- $TrVol$ is the daily change in the volume of transactions occurring on the network as (presumably) payment for goods and services.
- Ads is the log of the number of unique addresses participating in the Bitcoin network.

- HR is the log hash rate measured in GH/s.
- $GED(0,1,\kappa)$ is the generalized error distribution or generalized Gaussian distribution with mean 0, unit variance and shape parameter κ.

The volatility process outlined above is GARCH (1,1) which is by far the most common specification used in empirical finance. Very rarely is it appropriate to include additional lags of the innovation and variance term and we have no *a priori* reason to do so here. We have elected to have the exchange volume and transaction volume feature in the variance equation. The hypothesis being that for a large portion of Bitcoin's time in existence it was a very small market by absolute size and is still relatively small in comparison to other financial assets. As such it is reasonable to think there were times when the market was relatively illiquid and as such a single transaction may have made up a relatively large proportion of that day's total trading volume which could serve to "move the market" and induce large swings in the price.

Furthermore, we have assumed that the distribution of the error term is non-normal and rather follows a generalized error distribution. This distributional choice has the advantage that it enables us to account for possibly heavy tails in the distribution of the residuals by allowing for increased or decreased "peakedness". In the limiting case that the distribution is normal when κ=3.

Section B: Empirical Results and Diagnostic Checks

The above model was estimated via maximum likelihood for the first 2023 observations spanning dates 9-14-2011 through 3-29-2017; 365 days of observations were left out of the initial estimation to allow for out of sample testing. White's HAC standard errors were used to compute the relevant t-statistics for the estimated parameters. Bayes information criterion (BIC) and review of the literature were used in to determine the number of lags to include in the mean equation. Results are presented in Table 3.

Table 3: Estimation results for Bitcoin GARCH (1,1) process.

<u>Conditional Variance Dynamics</u>

Model Order: GARCH (1,1)

Mean Model: ARIMA (2,0,0) with no seasonality

Distribution Model: Generalized Error Distribution

Observations: 2023

Parameter	Estimate	Standard Error	t-statistic	p-value
Mean Model				
Constant	-.04134	.00021	-190.424	<.001
AR(1)	-.05166	.00653	-7.904	<.001
AR(2)	-.03185	.00322	-9.873	<.001
ExVol	.00263	.00032	8.126	<.001
TrVol	-.00295	.00072	-4.074	<.001
# of Addresses	.00467	.00015	310.363	<.001
Hash Rate (GH/s)	-.00119	.000014	-84.285	<.001
Variance Model				
Omega	.000042	.000003	14.063	<.001
Alpha (α)	.15283	.00768	19.901	<.001
Beta (β)	.82441	.000854	964.99	<.001
ExVol	.00035	.000023	15.235	<.001
TrVol	.000045	.000014	15.285	<.001
Shape (κ)	.8540	.03961	21.559	<.001

Information Criterion	Value
Akaike	-4.185
Bayes	-4.149
Hannan Quinn	-4.171

Our analysis will begin with a discussion of Bitcoin market efficiency, but first a distinction should be made. The Bitwise Asset Management study on inflationary transactions and their impact on Bitcoin markets, is probably the most credible study to date that seeks to draw conclusions about the relative efficiency of Bitcoin. The form of efficiency evaluated in that study is slightly different from that to follow. The Bitwise study, sought to determine whether Bitcoin is *arbitrage efficient*; that is, if the Bitcoin market is sufficiently integrated, then intermarket arbitrage should not be possible. A trader should not be able to purchase Bitcoin from one exchange and simultaneously sell it on another to generate a profit. Our concern here is with *information efficiency*; i.e. "how well is historical and present information incorporated into Bitcoin's price?"

Results indicate a high degree of statistical significance among all model regressors for both the mean and variance equation as well as the shape parameter. Proceeding in order, we observe that the coefficients for AR (1) and AR (2) are highly statistically significant. These coefficients represent the correlation of the BTC time series with lagged versions of itself at lengths of 1 and 2 days respectively. The statistically significant results would seem to imply that the Bitcoin market violates the weak-form of the Efficient Market Hypothesis. The EMH states that all historical price information is incorporated into current market prices. However, the magnitudes of the coefficient estimates are very small which indicates the effects may not be economically meaningful. Furthermore, a compelling argument can be made that the market is indeed weak-form efficient based on the submartingale argument advocated by Fama (1970); which posits a subtler standard for establishing weak form efficiency than necessitating the coefficients of the lagged series be zero.

Fama (1970) claims that a data generating process is submartingale if:

$$E(r_{t+1} | \phi_t) \geq 0$$

That is, the expected return next period (r_{t+1}) based on the information set (ϕ_t) available at time 't' is at least 0. If the statement holds with equality:

$$E(r_{t+1} | \phi_t) = 0$$

then the process is a martingale; the expected return in the next period given all information up to the present is '0' (i.e. a random walk).

The objective of a trader in a market is a produce a trading rule that is a submartingale. However, a submartingale in *returns* has an important empirical implication: a "buy or go to cash" trading rule based only on information about historical returns available at time 't' cannot produce greater profits than a simple "buy and hold" trading strategy. It follows that, in the presence of a submartingale in returns, a trading strategy based exclusively on historical returns data would not be able to generate excess returns.

Furthermore, the inability to find a submartingale trading strategy based on historical return data would also all fail to earn returns in excess of a simple "buy and hold" strategy; which (for all intents and purposes) would satisfy our definition of weak form efficiency.

Given the magnitude of the AR (1) and AR (2) coefficients of .05166 and .03185, respectively, the approximate variation explained by the lagged series accounts for only .36% of the total variation. At such a low level of explanatory power, it is unlikely that a trading strategy predicated on using only historical return data would not be exceedingly profitable (particularly in the presence of any trading costs or market frictions). As such, we conclude that the Bitcoin market does exhibit weak form efficiency.

Results in the mean equation suggest that both supply and demand factors are important for understanding how the price of Bitcoin has developed since inception. The interplay between the model variables is interesting and may give us some clues as to how Bitcoin is being used. The coefficient corresponding to the number of unique addresses is positive and extremely significant. This implies that increased adoption of Bitcoin has had a positive impact on price returns as the demand for Bitcoin has increased and confirms our expectations *a priori*. This is observed in conjunction with a positive and significant coefficient for the change in exchange volume and a negative, significant coefficient for the change in transaction volume. This may confirm the hypothesis that Bitcoin is being viewed more as a speculative asset than an actual currency as exchange volume has a positive impact on price returns (confirming *a priori* expectations), but network volume has a negative impact which does not align with what we had anticipated.

The interpretation of the coefficient for the hash rate offers particularly interesting results. The high degree of significance confirms our expectations that this variable plays a critical role in supporting the Bitcoin ecosystem by ensuring that transactions can be processed. The negative coefficient could be interpreted as a supply side effect where when the block reward is distributed to miners, the miners immediately turn around and sell the coins in the market thereby continually injecting fresh liquidity which increases not only the number of coins in circulation (which is a given based on how the network operates), but the number of coins trading thus inducing a negative price return effect through an increase in market supply which must be soaked up. This could be an ominous sign for the longevity for the BTC network as rather than holding onto their block reward and pursuing long run price appreciation the miners (who presumably have the most knowledge regarding BTC and its possible future) are seen simply cashing out for more stable and immediately usable currency.

Turning to the volatility equation we see some reassuring results. Again, all coefficients are significant. More importantly, the coefficients alpha and beta sum to comfortably less than one; $\alpha + \beta = .1528 + .8244 = .9772 < 1$. This is very good as it shows that the volatility equation is stationary and not explosive. A result greater than 1 would indicate the presence of a unit root in the volatility equation implying that periods of high volatility are following by periods of even higher volatility about which we could say very little. The positive coefficients on the exchange and transaction variables are significant, but relatively small in magnitude implying that periods of heavy trading increases volatility, but that the effect is not terribly substantial.

Finally, the estimate of the shape parameter is highly significant and substantially less than 3. This implies that the residuals are not adequately modeled by a normal distribution and that the distribution exhibits a significant degree of "peakedness".

Next, we run a series of diagnostic checks on the above model to assess how well it is specified and its usefulness in application. First, we test for an adequately fit ARCH process using the ARCH LM test statistic proposed in Li and Mak (1994)[33]. Li and Mak were able to show that the more familiar Box-Pierce ***Q***[34] and Ljung-Box[35] test statistics that had been adapted to detect autocorrelation in the variance were not asymptotically distributed chi-squared when constructed with squared residuals. They propose a statistic for a fitted GARCH process based on the autocorrelation function of the standardized squared residuals constructed as follows:

$$\hat{r}_k\left(\frac{\hat{\varepsilon}_t^2}{\sigma_t}\right) = \frac{\sum_{t=t-k}^{n}(\frac{\varepsilon_t^2}{\sigma_t} - \bar{\varepsilon})(\frac{\varepsilon_{t-k}^2}{\sigma_{t-k}} - \bar{\varepsilon})}{\sum_{t=1}^{n}(\frac{\varepsilon_t^2}{\sigma_t} - \bar{\varepsilon})^2}$$

The test statistic can then be constructed as follows:

$$L(p,m) = n \sum_{k=p+1}^{m} \hat{r}_k^2$$

Where:

- p is the number of ARCH parameters fit to the process.
- k is the beginning lag length which must be at least 2 because at least 1 ARCH parameter must be fit to the model.
- m is the maximum lag length for which we are interested if autocorrelation exists.

The test statistic has a null hypothesis of an adequately fit ARCH model (i.e. we have the correct number of p's) and has been shown to be asymptotically distributed chi-squared with *m-1* degrees if freedom. Table 4 presents the result of this test. The results indicate that ARCH effects have been sufficiently eliminated by the model.

Table 4: Results of Li Mak ARCH LM Test

Lag Length(m)	Test Statistic	d.o.f	p-value
2	.642	1	.4229
3	1.772	2	.4123
4	2.129	3	.5459
5	2.236	4	.6924
7	2.804	6	.8330
9	2.821	8	.9451

We assess the goodness of fit of the distributional assumption for the model residuals using both visual and analytic methods. Figure 4 presents a QQ-plot for the theoretical quantiles of the generalized error distribution against the standardized residuals of the model. Figure 5 plots a histogram of the model residuals and overlays a normal distribution and GED using the parameters estimated from the model. The residuals in the QQ-plot fall reasonably close to the straight line, but the plot suggests the continued presence of heavy tails. The plot of the theoretical histogram does confirm a significant departure from normality and it appears that the generalized error distribution does fit the data somewhat more well.

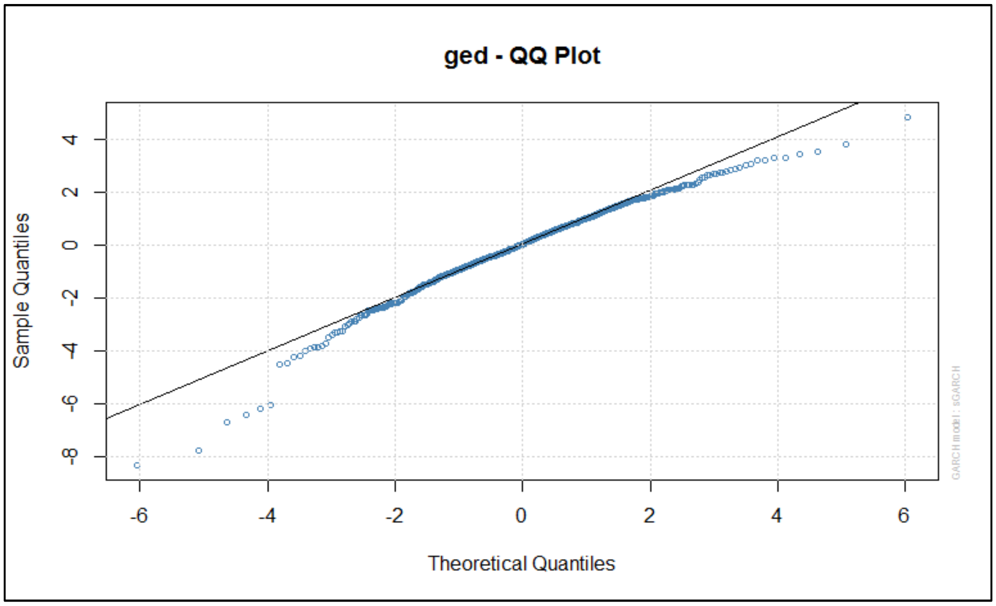

Figure 4: QQ-plot for standardized residuals against GED Distribution

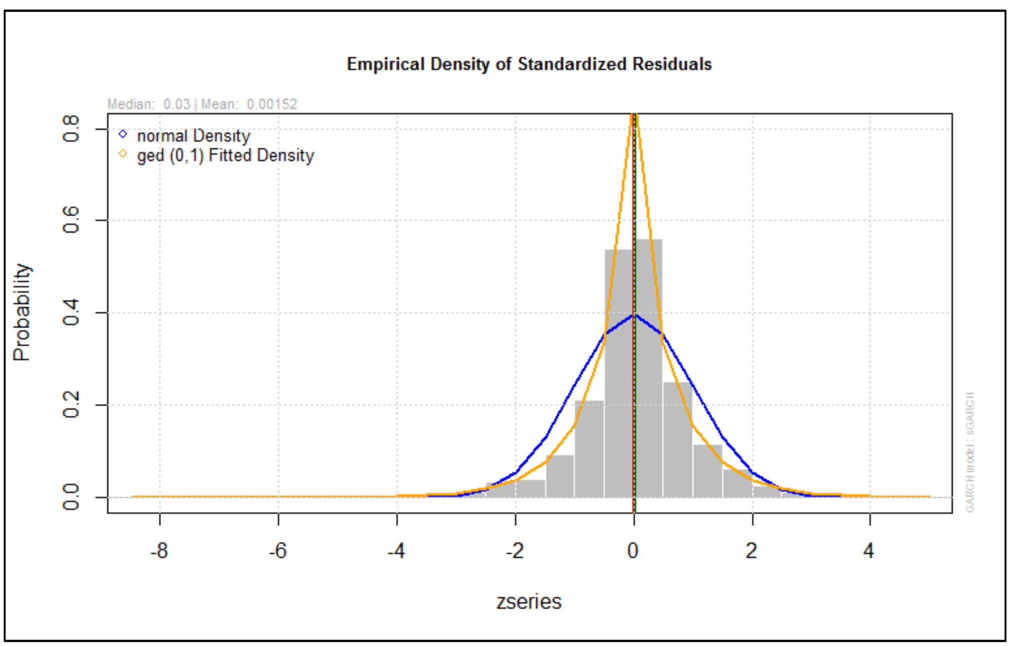

Figure 5: Empirical Histogram of Standardized Residuals

These visual methods suggest that the distributional assumption is probably sufficient, but a more formal test is needed. We use an augmented version of Pearson's Goodness of Fit Test to test the empirical distribution of the standardized residuals with the theoretical ones predicted by the GED density. The implementation of the test is from Palm (1996)[36] who proposed a goodness of fit test with a null hypothesis of adequate fit that is robust in the presence of heteroskedasticity and is asymptotically distributed chi-squared. The test was repeated with varying bin sizes. The results are presented in Table 5.

Table 5: Goodness of Fit Test for model residuals (d.f.=n-1)

Bin Group (n)	Statistic	p-value
20	16.79	.6039
30	35.79	.1798
40	46.38	.1941
50	54.68	.2677

The test results indicate a strong degree of fit for the standardized residuals given the empirical distribution.

In the above model we have assumed that volatility is symmetric; that is, the ARCH effect introduced by including the previously observed innovation has the same effect on the volatility regardless of the sign. A common feature of financial time series models is the leverage effect or the tendency for volatility to increase more in the presence of negative innovations (i.e. bad news) then in the presence of positive innovations (i.e. good news). Engle and Ng (1993)[37] propose the "news impact" or sign bias test to assess the leverage effect. The presence of leverage in the standardized residuals may indicate that the proposed GARCH model is mispecified. Engle and Ng suggest performing the following auxiliary regression on the standardized residuals and the lagged positive and negative shocks:

$$\hat{z}_t^2 = \lambda_0 + \lambda_1 I_{\hat{\varepsilon}_{t-1<0}} + \lambda_2 I_{\hat{\varepsilon}_{t-1<0}} \hat{\varepsilon}_{t-1} + \lambda_3 I_{\hat{\varepsilon}_{t-1\geq 0}} \hat{\varepsilon}_{t-1} + u_t$$

Where:

- $\hat{z}_t^2$ is the squared standardized residual at time t.
- $\hat{\varepsilon}_{t-1}$ is the lagged residual from the GARCH process.
- I is a dummy variable that takes the value of 1 when the innovation is negative and 0 otherwise.

Sign bias for the model can be evaluated for each sign individually or jointly using *t* and LM tests respectively. A significant result for λ_1 would indicate general sign bias; that is, an asymmetric effect in volatility resulting from positive and negative innovations regardless of size. Inclusion of λ_2 and λ_3 enable us to control for the size of the effect. In this way we can evaluate if the relative size of the innovation and sign also has an effect; the hypothesis being that large negative/positive innovations (i.e. very bad or very good news) induce greater changes in volatility than if we just consider sign. For the hypothesis H_0: $\lambda_1 = \lambda_2 = \lambda_3 = 0$ we compute the following test statistic which Engle and Ng has shown to be asymptotically distributed chi-squared with 3 d.f.

$$n * R_{asy}^2 \sim \chi_3^2$$

allows us to evaluate if the impacts are jointly significant. Results of the test are presented in Table 6. Figure 6 depicts the News Impact Curve which provides a visual representation of how the size and sign of the innovations affects volatility.

Table 6: Results Sign Bias Test of model residuals

Test Type	Statistic	p-value
Sign Bias (λ_1)	.1058	.9158
Negative Sign Bias (λ_2)	1.4939	.1354
Positive Sign Bias (λ_3)	.3308	.7408
Joint Effect	2.7047	.4394

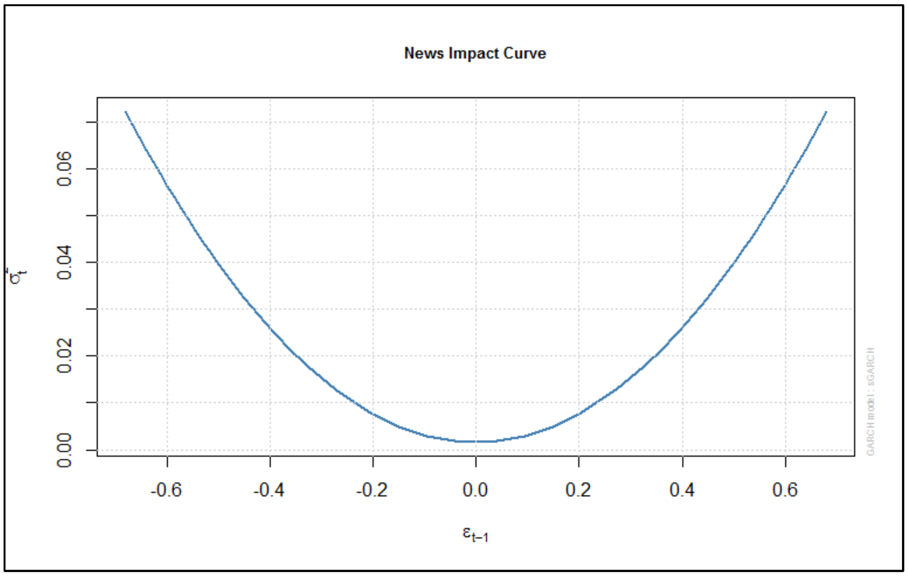

Figure 6: News Impact Curve

Test results do not the indicate the presence of asymmetric sign bias which suggests that Bitcoin price returns respond similarly to both positive and negative news. Additionally, the news impact curve shows a smooth symmetric impact to volatility; if sign bias were present then we would expect the curve to be tilted or exaggerated about the 0-axis.

So far, diagnostic tests have suggested that this model is well specified, but ideally the model should have applicability in a forecasting setting. A well specified model should be able to quickly and accurately incorporate new information about Bitcoin's return process and capture changes in volatility. The ability to forecast can be evaluated using Value at Risk (VaR) and expected shortfall. Value at risk is a technique commonly applied in finance and banking that measures the magnitude of loss on an investment at a specified probability. It estimates how much an investment might lose given a specific probability and thereby enables investors and managers to gauge the amount of assets needed to cover the liability in a "worst case" scenario. Similarly, expected shortfall (ES) is the average loss that we would expect to observe given that the value at risk threshold has been exceeded. As such, both the VaR and ES estimates will vary based on the specified probability with higher probabilities producing more conservative estimates. We will apply a couple of different formulations of these tests and examine how our model performs in both in and out of sample testing.

The most common way to test a VaR model is to count the number of exceedances i.e. the number of days during the holding period where the portfolio VaR was exceeded given a certain level of confidence. For example, if we observe an interval of 100 trading days and establish a 99% VaR confidence interval for the daily returns of the portfolio, we would expect that threshold to be exceeded only once over the period. During backtesting we can statistically determine whether the frequency of exceedances over the time interval is sufficiently in line with our expectations. For our purposes, the VaR test described below will be implemented using the conditional volatilities derived from the GARCH model but could theoretically also be performed using the unconditional standard deviation for the analysis period.

Among the first of the unconditional coverage tests was Kupiec's (1995) proportion of failures (POF) test.[38] The test is referred to as an unconditional test because it assumes that the exceedances are independent. That is, observing an exceedance yesterday does not change the probability of observing an exceedance today. Kupiec proposed that the number of exceedances follows a binomial distribution:

$$f(x) = \binom{t}{x} p^x (1-p)^{t-x}$$

Where:

- p is the expected failure rate (ex. 1% for a 99% confidence level)
- x is the observed exceedance
- t is the number of observations in the sample

The null hypothesis of the test is that $p = \hat{p} = \frac{x}{t}$; that is, the observed proportion of exceedances is equal to the expected proportion. Kupiec proposes the following likelihood ratio test statistic:

$$LR_{uc} = -2\ln\left(\frac{(1-p)^{t-x}p^x}{\left(1-\frac{x}{t}\right)^{t-x}\left(\frac{x}{t}\right)^x}\right) \sim \chi_1^2$$

He demonstrates that the test statistic of a correctly specified model follows a χ^2 distribution with 1 degree of freedom.

Christoffersen et al. (2004) argued that counting the number of exceedances was insufficient for developing a robust test and that the test may fail to reject a model that produces clustered violations.[39] In his view, an ideal model would return value at risk exceedances that appear random and don't cluster i.e. a violation on one day is followed by a violation on the following day. Christoffersen argues that this is a stronger test of model adequacy and offers more robustness in application as it evaluates if observed VaR violations are independent of one another. This method is viewed as crucially important in the banking sector where consecutive days of severe capital losses could lead a bank to fail very quickly; the consequences of which are rather unsavory. While this is not the motivation here this test can still be viewed as good indicator of a well specified model.

The construction of the test statistic involves defining a dummy variable that takes the value of 1 if the value at risk is exceeded and 0 otherwise:

$$I_t = \begin{Bmatrix} 1 \ if \ r_t < VaR_t \\ 0 \ otherwise \end{Bmatrix}$$

Christoffersen proceeds to define the variable n_{ij} as the number of days when event *j* was observed after event *i*. For example, n_{10} would indicate a day when a violation was not observed following a day when a violation was observed while n_{00} would imply that a violation was not observed today or during the previous day, etc. Christoffersen presents this in a two-way contingency table as follows:

	$I_{t-1} = 0$	$I_{t-1} = 1$	
$I_t = 0$	n_{00}	n_{10}	$n_{00} + n_{10}$
$I_t = 1$	n_{01}	n_{11}	$n_{01} + n_{11}$
	$n_{00} + n_{01}$	$n_{10} + n_{11}$	N

Furthermore, let $\pi_0 = \frac{n_{01}}{n_{00}+n_{01}}$ and $\pi_1 = \frac{n_{11}}{n_{10}+n_{11}}$. Under the null hypothesis of a correctly identified model then $\pi_0 = \pi_1$; that is, the probability that we observe a VaR violation does not depend on whether we observed such an outcome on the previous day. Consequently, this would also imply that $\pi_0 = \pi_1 = p$; that the probability of observing an exceedance following a day where an exceedance was observed/not observed is equal to the probability that we observe an exceedance at all. The test statistic, which Christoffersen has shown to be distributed chi-squared with 1 degree of freedom. can then be formed as follows:

$$LR_{Ind} = -2\ln\left(\frac{(1-p)^{n_{00}-n_{10}}p^{n_{01}-n_{11}}}{(1-\pi_0)^{n_{00}}\pi_0^{n_{01}}(1-\pi_1)^{n_{10}}\pi_1^{n_{11}}}\right) \sim \chi_1^2$$

We present the results for these tests in two parts. Table 7A depicts the backtested results based on the initial 2023 observations used to develop the GARCH model, the fitted values of that model and the volatility. The table presents the results of both Kupiec's unconditional test and Chistoffersen's independence test at the 5% and 1% levels. Figures 7 shows the observed time series in Blue overlaid with VaR confidence bands The Red line depicts the forecasted value at risk at 95% confidence while the Black line traces the value at risk at 99% confidence.

Table 7A: Kupiec's Unconditional Coverage Test for In-Sample Fit

Alpha	5%	1%
Expected Exceedance	101	20
Observed Exceedance	100	28
χ_1^2 Critical Value	3.841	3.841
χ_1^2 Test Statistic	.0138	2.692
p-value	.906	.101
Decision	"Fail to Reject Null"	"Fail to Reject Null"

Table 7B: Christoffersen's Test of Independence for In-Sample Fit

Alpha	5%	1%
χ_1^2 Critical Value (5% level)	3.841	3.841
χ_1^2 Test Statistic	10.456	3.53
p-value	.013	.0602
Decision	"Reject Null"	"Fail to Reject Null"

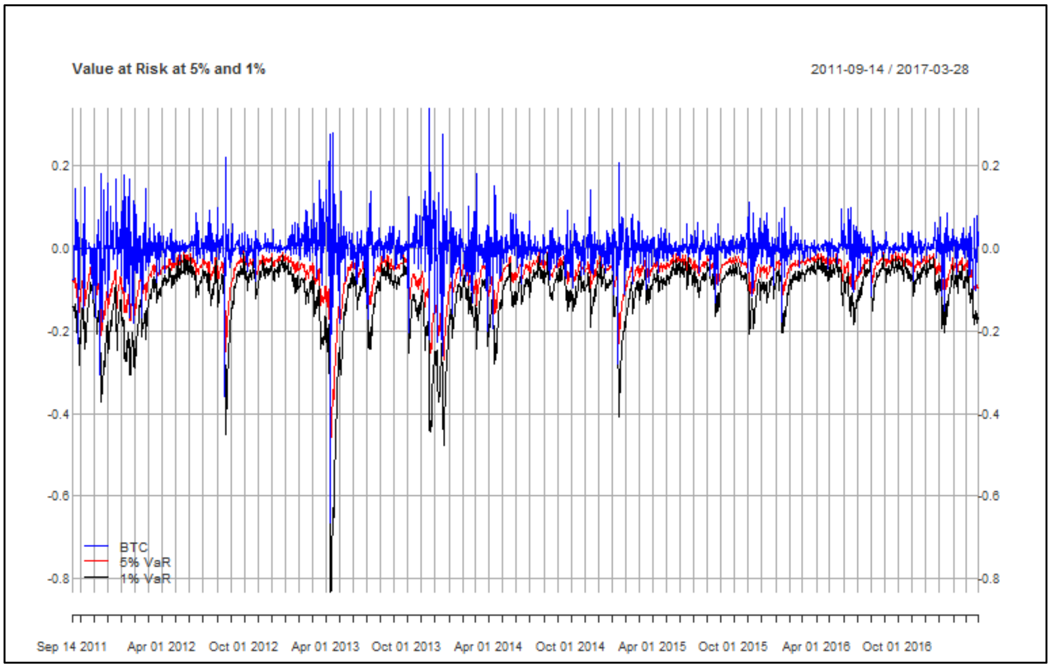

Figure 7: Bitcoin In-Sample Return Plot with 5% and 1% VaR bands

Examination of the results for in sample fit offer a mixed perspective. Using Kupiec's test we fail to reject the null hypothesis of the correctly observed number of value at risk violations at both a 5% and 1% level. In the 5% case, the rejection is quite strong, and we are almost spot on in terms of predicted vs. observed violations. The results at the 1% level are not quite a strong and could be argued as perhaps borderline weakly significant, but for our purposes the model still passes this test comfortably.

The results from Christoffersen's independence test are somewhat spottier. At a 5% level for value at risk the hypothesis of independent exceedances is rejected strongly suggesting that it is more likely to observe a 5% violation on the day following a 5% violation then we would otherwise anticipate. We fail to reject the model at a 1% level, but just barely. While a p-value of 6.2% is not sufficient to reject the hypothesis of independent 1% exceedances it is uncomfortably close and a less strict standard for failure (i.e. 10%) would lead us to reject independence.

Examination of expected shortfall (ES) can also be used to evaluate the in-sample fit of the model. Expected shortfall is designed to measure the loss that can be expected given that the VaR threshold has been exceeded at a pre-specified probability. Specifically, the ES is the average of the worst 100(1-α) % of losses:

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_p dp$$

Where α is the confidence level (i.e. 90%, 95%, etc.) and q_p is the quantile distribution.

To evaluate in-sample fit we will evaluate ES using a 95% confidence level; corresponding to a 95% VaR. From our results on VaR above we observed that over the in-sample period of September 14th, 2011 to March 28th, 2017, 5% VaR was exceeded 100 times which will provide a reasonably sized sample to conduct the analysis. A histogram of the density of observed losses of Bitcoin exceeding the 5% threshold is presented in Figure 8; the histogram and loss values have been presented as absolute values for expositional convenience. The histogram represents the left tail of the Bitcoin return distribution and as such is naturally skewed.

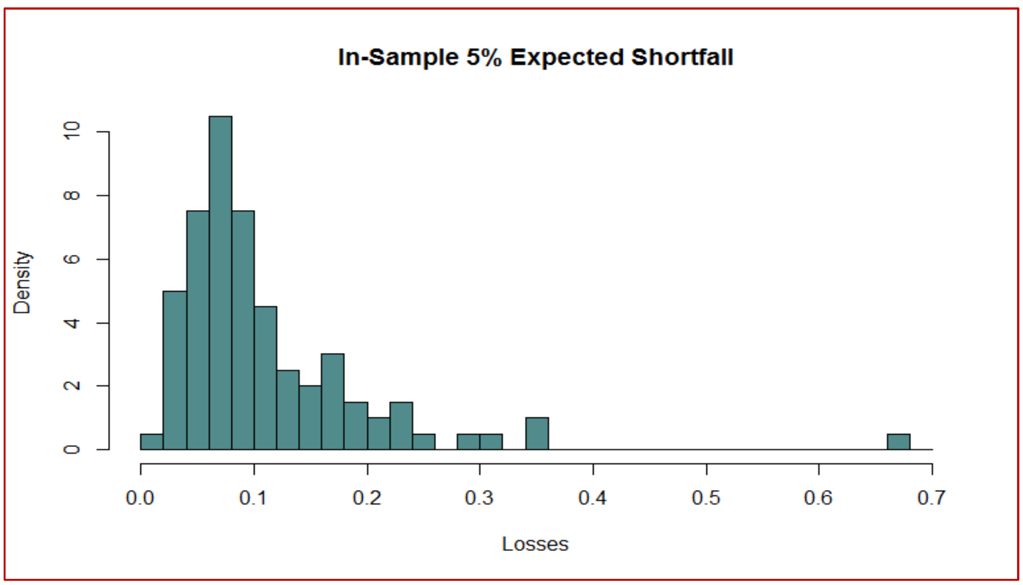

Figure 8: Bitcoin Expected Shortfall at 5%

We can construct a theoretical distribution for the expected shortfall for our model using the conditional mean and conditional volatility forecasts derived from the GARCH model as follow:

$$\widehat{ES_{\alpha,t}} = \hat{r}_{BTC,t} + \hat{\sigma}_{BTC,t} * \frac{1}{1-\alpha}\int_\alpha^1 q_p dp$$

Because expected shortfall is essentially the mean of the loss distribution, we can apply a simple t-test to the estimated ES and the observed ES to determine whether the two are equivalent. The results of the difference in means test is presented in Table 8.

$$H_0\text{: } \widehat{ES_{\alpha,t}} - ES_{\alpha,t} = 0\text{; the means are the same.}$$

$$H_a\text{: } \widehat{ES_{\alpha,t}} - ES_{\alpha,t} \neq 0\text{; the means are different.}$$

Table 8: Expected Shortfall Test for Difference in Means

	Observed Shortfall	Theoretical Shortfall
Mean	-10.99%	-9.48%
Standard Error	8.92%	5.64%
Degrees of Freedom	167.35	
Alpha	.05	
Test Statistic	-1.4303	
p-value	.07725	
Conclusion	Fail to reject the null hypothesis of equal means.	

The true test of any time series model is its ability to forecast. The time period we used to calibrate the model spanned September 14, 2011 to March 28, 2017. The last year we have left out to evaluate the performance of the model out of sample. This period is of interest because it includes what many believe was the peak and subsequent pop of the Bitcoin bubble. On March 29, 2017 Bitcoin's price was $1,035.96. At its end of day peak on December 16, 2017 the price was $19,187.78; an increase of over 1,700% in a little less than 9 months' time. This meteoric rise was quickly followed by an equally impressive collapse of -58% to a price of $8,039.86 on March 28, 2018. The point of this anecdote is to highlight the extent of the price swings that Bitcoin has experienced over the past year alone and lend some context to the test results to follow.

The forecast for the Bitcoin price return series was constructed using rolling 1-step ahead forecasts. Using this methodology, a forecast model is initially specified using *t* observations; in our case, *t* is equal to the initial 2023 observations used to specify the GARCH model so the initial forecasting model is identical to what has already been presented. A single forecast is then made for period *t+1*. The forecast for period *t+1* is then compared to the true observed value at *t+1*, the residual is calculated and that information is the fed back into the GARCH model to develop a forecast for period *t+2*'s return and volatility. Furthermore, for the first 50 observations we assumed that the parameters of the model were stable; that is, the only thing that changed from period to period was the new information being incorporated after each additional forecast not the model parameters themselves. After every 50 of these forecasts the model was reestimated to allow the parameters to change in the presence of new information. The refit was conducted using the entire time series up to the last forecast so the first refit used *t+50* observations, the second *t+100*, etc. This lends the model a dynamic quality that should theoretically increase the forecasting accuracy; thus, the rolling forecast from March 29, 2017 to March 28, 2018 includes 7 model refits.

Kupiec's and Christoffersen's tests were then implemented in the same way as with the in-sample procedure to assess the model's forecasting capabilities; the results for which are shown in Tables 9A and 9B. Figure 9 provides the visual representation of the forecasted VaR against the observed data.

Table 9A: Kupiec's Unconditional Coverage Test for Out-Sample Fit

Alpha	5%	1%
Expected Exceedance	18.2	3.6
Observed Exceedance	27	4
χ_1^2 Critical Value	3.841	3.841
χ_1^2 Test Statistic	3.873	.033
p-value	.049	.856
Decision	"Reject Null"	"Fail to Reject Null"

Table 9B: Christoffersen's Test of Independence for Out-Sample Fit

Alpha	5%	1%
χ_1^2 Critical Value (5% level)	3.841	3.841
χ_1^2 Test Statistic	1.88	1.971
p-value	.17	.160
Decision	"Fail to Reject Null"	"Fail to Reject Null"

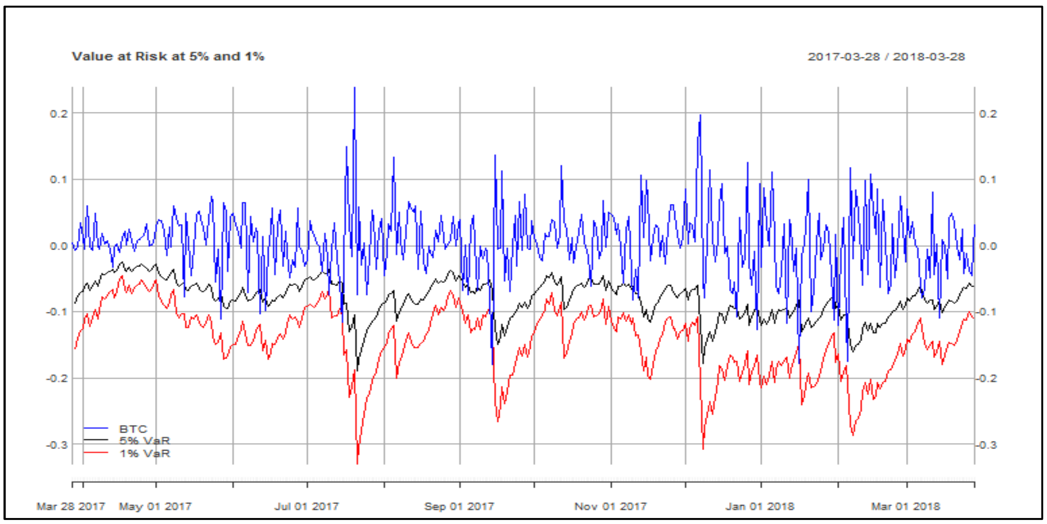

Figure 9: Bitcoin Out-Sample Return Plot with 5% and 1% VaR bands

Results from Kupiec's test are somewhat mixed. At a 5% VaR we would practically anticipate 18 violations, but 27 observed violations are 50% higher than expected; an excessively high number that ultimately leads us to reject the null hypothesis of correct exceedances. At a 1% level the model performs quite well as the number predicted vs. observed exceedances is very close. Christoffersen's test also provides some encouraging results as we fail to reject the null hypothesis of independent exceedances at 5% and 1% levels. Use of the graph is also informative in this case as we can almost simply count the number of times a Bitcoin return fell below one of the VaR bands and if those exceedances cluster or show unusual patterns. In this case the violations do appear random and no violation is observed that is significantly lower than the 1% band.

As mentioned previously, a functional currency will meet 3 criteria: it will serve as 1) a medium of exchange, 2) a store of value, and 3) a unit of account. There must be a certain level of stability and predictability associated with a currency such that consumers can reliability transact with it and hold cash while knowing that its value will be preserved across time. The price return process enables us to determine how tenable these objectives are for Bitcoin. It is crucial to understand the risks and be able to reasonably forecast movements in the price of a currency. By understanding these underlying actions, it becomes possible to determine possible/appropriate use cases for BTC and other cryptocurrencies. As such, the Value at Risk and Expected Shortfall tests are helpful for evaluating forecasting capabilities and the probability and scale of large losses which will be crucial if institutions are to develop real use cases for Bitcoin such as eligibility to serve as collateral or to issue debt denominated in the currency.

The results of the VaR and ES tests are challenging in that they do not strongly confirm or reject our expectations. The unconditional coverage VaR tests suggest that we can accurately measure losses associated with low probability events. This is essential if a financial institution is going to extend a loan collateralized by BTC. However, the VaR-independent exceedance tests suggest that large daily losses are followed by more large losses the following day. This is problematic if BTC is to serve as eligible collateral as capital calls would likely be quite frequency and the advance the institutions would be willing to make would likely be so small as to be useless. This is to say nothing of more complex dependencies in the data (beyond just 1 day) that we have not investigated.

The ES shortfall test presents a similar problem. Even if we cannot reliably say that VaR exceedances are independent, the issue may be mitigated to a certain degree if we can consistently measure the size of the losses. For example, a series of 1% daily losses may be undesirable for a bank but need not be disastrous if the bank can measure this possibility and ensure it has sufficient capital on hand. With respect to Bitcoin, the ES test result is affirmative (with a p-value of approximately 7%) that this can be predicted but is softer than might be desired and the size of the losses (at nearly 10% on average) is probably unacceptable for even the riskiest institutions.

Deployment of specific use cases around Bitcoin and its properties as a currency will be required to ensure stability and longevity of the currency. However, these tests cast considerable doubt about the possibly of Bitcoin gaining wide-spread legitimacy and acceptance.

Section C: Concluding Remarks on Objective 1

Objective 1 concerned the estimation and evaluation of Bitcoin's return and volatility process. Considering that Bitcoin is 2-3x more volatile than traditional financial assets fitting a model that is dynamic and can accurately forecast is necessary if BTC is to experience greater adoption by institutions and the public. Bitcoin is a unique financial asset because it trades 24/7/365 and as such it is appropriate to model the cryptocurrency based on qualities that could be continuously observed. To do this, we proposed a GARCH (1,1) model that incorporates both demand and supply characteristics of Bitcoin such as: exchange volume, transaction volume, the number of unique addresses and the hash-rate. The results from the modeling procedure indicated that these characteristics play an integral role in Bitcoin's development. We further subjected the model to a series of diagnostic tests which sought to assess the adequacy of Bitcoin's statistical properties and might allow us to conclude that the model is well specified. Based on these tests we determined that the GARCH (1,1) process fits well. Finally, we investigated Bitcoin's value at risk. If our model does, in fact, describe Bitcoin's return process then we should be able to accurately quantify and forecast Bitcoin's volatility and capture the inherent risk. Test results suggest that our model can forecast Bitcoin's volatility reasonably well and that we were adequately able to predict the number of VaR violations.

CHAPTER 6: MACRO-FINANCIAL ANALYSIS

Section A: Econometric Approach

The focus of Objective 1 was to model and evaluate Bitcoin's daily return and volatility process. The hypothesis being that Bitcoin's process is driven by variables that are unique to Bitcoin and therefore the analysis was conducted using principally supply and demand variables for Bitcoin. It has been suggested however that Bitcoin displays a significant relationship with many common macroeconomic and financial variables and that these variables play an integral role in BTC price development ((Van Wijk (2013)[11] and Dyhrberg (2015)[14]). The motivation for Objective 2 then is to assess Bitcoin's exposure to macro-financial variables and quantify the risk those variables contribute to Bitcoin's overall risk profile.

Analyzing Bitcoin's relationship to traditional macro-financial variables presents a unique set of challenges as previously discussed in Section IV. Considering these challenges, on the outset we harbor some doubts about the degree to which these variables influence Bitcoin. The variables we propose are familiar in financial time series analysis and were selected to broadly represent the dynamics of important, related markets. Included are the following:

- *Federal Funds Rate (FF)*: Arguably the most widely quoted and followed rate in the world, the Fed Funds Rate is the interest rate that U.S banks charge one another on overnight, uncollateralized loans and often serves as the basis for the construction of other interest rates such as the prime rate. The Federal Reserve sets the target for this rate and it can be used as a proxy for the monetary policy stance of the Federal Reserve.
- *Change US Dollar-Japanese Yen Exchange Rate (USD-JPY):* 2nd most heavily traded currency pair in the world. Both the Yen and Dollar are recognized as reserve currencies. The exchange rate is very stable and may be used as a gauge of the health of the U.S and Japanese economies and worldwide trade.

- *Change US Dollar-Euro Exchange Rate (USD-EUR):* Most heavily traded currency pair in the world. Both the Euro and Dollar are recognized as reserve currencies. The exchange rate is very stable and may be used as a gauge of the health of the U.S and Eurozone economies and worldwide trade.
- *Return on S&P 500 (SPX):* The S&P 500 is broadly considered to be a good gauge of the health of financial markets in the U.S and across the world. It is often used a proxy for the market portfolio and features prominently in many applied financial models (CAPM, etc.).

The selected variables draw on those used previously by Van Wijk (2013)[11], Dyhrberg (2015)[14], and Ciaian, Rajcainova and Kancs (2015)[12]. Table 10A presents the descriptive statistics for these additional model variables and Bitcoin along with a correlation matrix (Table 10B) for the period January 6, 2012 through March 23, 2018.

Table 10A: Descriptive Statistics and Correlation Matrix

Statistic	BTC-USD Ex Rate	FF	USD-JPY	USD-EUR	SPX
Observations	325	325	325	325	325
Mean	.02287	.00352	.0009	-.0001	.0022
Median	.01678	.0014	.0007	.0004	.0031
Minimum	-.5937	.0006	-.0421	-.0363	-.0614
Maximum	.5398	.0168	.0483	.0288	.0446
Std. Deviation	.1288	.00397	.0136	.0117	.0164
Skewness	.068	1.562	-.0823	-.2282	-.6979
Kurtosis	3.148	1.131	.5644	.1238	1.7686

Table 10B: Correlation Matrix

	BTC	FF	USD-JPY	USD-EUR	SPX
BTC	1.000				
FF	.019	1.000			
USD-JPY	-.008	-.108	1.000		
USD-EUR	.027	.103	-.323	1.000	
SPX	.031	-.026	.332	-.0006	1.000

Examination of the descriptive statistics shows that Bitcoin is substantially more volatile than the other financial assets (USD-JPY, USD-EUR, and SPX), but that risk taking appears to have been justified so far as the weekly return is some 10x greater than the S&P 500. The correlation matrix depicts a very low degree of correlation between Bitcoin and the other financial assets which is in line with the results of Bouri et al. (2016)[16] who suggested that Bitcoin can serve as a reasonable portfolio diversifier because of its low correlation with other asset classes. As expected, the correlation between Bitcoin returns and the Fed Funds rate is quite low. Bitcoin is not controlled by any central authority, does not exhibit an interest rate regime and (for much of its existence) has been disconnected from the world economy at large. These factors taken together would suggest that Bitcoin returns should not be substantially impacted by Fed policy.

The proposed GARCH draws on the results obtained from Objective 1 and hence includes a mix of both demand-supply variables and macro financial factors in the mean equation. The model has been simplified to some degree with respect to the variance equation and the distributional assumption; specifically, the change in exchange trading volume and change in transaction volume variables have been dropped. Furthermore, the distributional assumption made for maximizing the likelihood function has been changed from the Generalized Error Distribution to a Student t distribution. The reasons for doing so are technical. The inclusion of the exchange volume and transaction volume variables created a convergence problem in **R** when maximizing the log-likelihood function under the GED assumption. To try and overcome this problem the model was estimated again in Stata, but a similar error was encountered. The source of these issues is somewhat uncertain, but likely has do to with the implementation of the maximization procedure, the convergence tolerance or both. As such, the distribution was simplified to Student t which is still a common choice for financial time series models and is well documented in the literature by Bollerslev (1986)[23]; who, at this point, we defer to as the authority on such matters. These changes appear to have rectified the convergence problem, allowed us to account for fat-tails in the BTC return process and, ultimately, produced a model that enables us to evaluate our primary concern.

However, it is also possible that the model is misspecified and that the convergence problems arose simply because the model is wrong. While we do not believe this to be the case the reader should interpret the results with this possibly in mind

The model we propose is as follows:

$$r_{BTC,t} = \beta_0 + \theta_1 r_{BTC,t-1} + \beta_1 ExVol_t + \beta_2 TrVol_t + \beta_3 Ads_t + \beta_4 HR_t + \beta_5 FF_t + \beta_6 JPY_t + \beta_7 EUR_t + \beta_8 SPX_t + a_t$$

where

$$a_t = \varepsilon_t \sigma_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_{t-1} \sigma_{t-1}^2$$

and

$\varepsilon_t \sim t(0,1,m)$ where m is the degrees of freedom

Where:

- FF_t is the log of the Federal Funds rate at time *t.*
- JPY_t is the log change in the USD-Yen exchange rate from time *t-1* to *t.*
- EUR_t is the log change in the USD-Euro exchange rate from time *t-1* to *t.*
- SPX_t is the log return of the S&P 500 index from time *t-1* to *t.*
- And the other model variables are consistent with those previously defined.

Section B: Empirical Results and Diagnostic Checks

The above model was estimated via maximum likelihood for weekly observations spanning dates January 6, 2012 through 3-23-2018; a total of 325 observations. White's HAC standard errors were used to compute the relevant t-statistics for the estimated parameters. Bayes information criterion (BIC) and review of the literature were used in to determine the number of lags to include in the mean equation. Results are presented in Table 11.

Table 11: Estimation Results for Bitcoin GARCH (1,1) process

Conditional Variance Dynamics

Model Order: GARCH (1,1)

Mean Model: ARIMA (1,0,0) with no seasonality

Distribution Model: Student's t Distribution

Observations: 325

Log Likelihood: 277.91

Parameter	Estimate	Standard Error	t-statistic	p-value
Mean Model				
Constant	-.8003	.0192	-41.57	<.001***
AR (1)	.0009	.0623	.0157	.9874
Ex. Vol	.0002	.0090	.0315	.9748
Trans. Vol	.0169	.0257	.6564	.5115
# of Addresses	.0917	.0012	71.425	<.001***
Hash Rate (GH/s)	-.0246	.0018	-13.214	<.001***
FF	.0106	.0101	1.047	.2947
JPY	.1593	.4787	.3328	.7392
EUR	.1554	.4632	.3355	.7372
SPX	-.0375	.3823	.0982	.9217
Variance Model				
Omega	.0011	.0007	1.506	.131
Alpha (α)	.4098	.1255	3.263	<.001***
Beta (β)	.5891	.0742	7.939	<.001***
Shape (*m*)	5.136	1.684	3.049	.002***

Information Criterion	Value
Akaike	-1.624
Bayes	-1.461
Hannan Quinn	-1.559

Below we present a brief set of diagnostic checks conducted on the estimated model to assess fit and distributional adequacy. The tests are identical in implementation are those discussed in Objective 1.

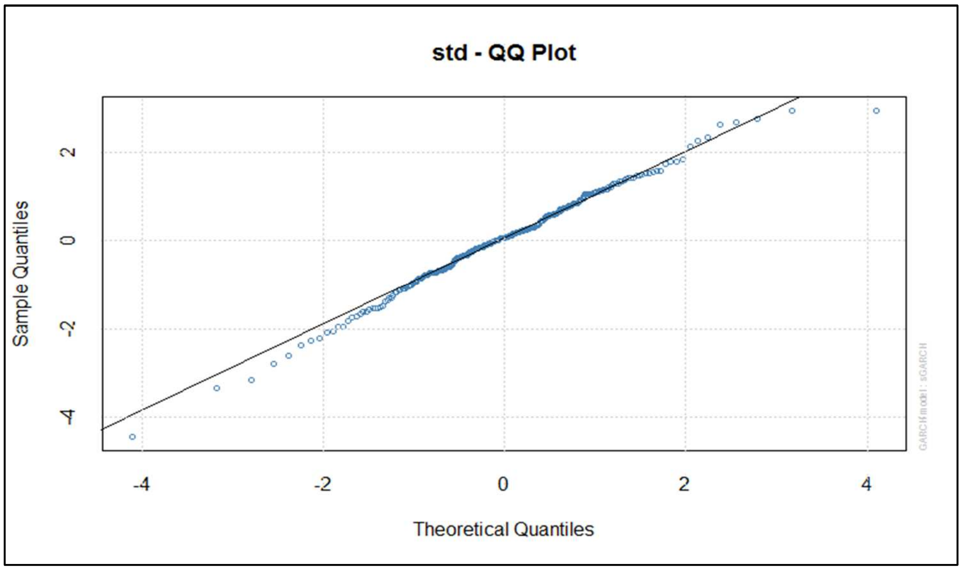

Figure 10: QQ-Plot of Standardized Residuals for Student’s t-Distribution

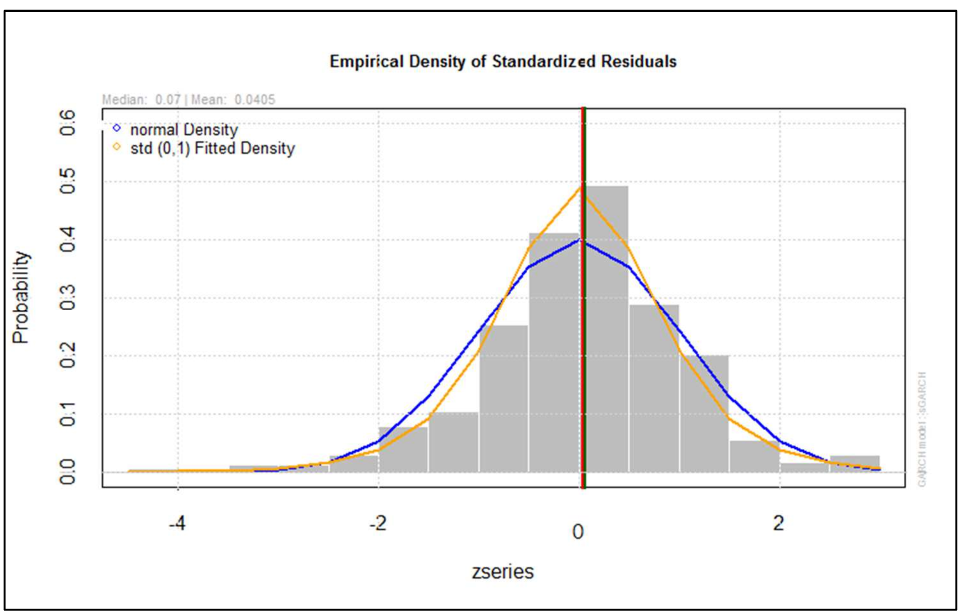

Figure 11: Empirical Distribution of Standardized Residuals

Visual inspection of the QQ-plot of the model residuals against the theoretical quantiles corresponding to the Student's t distribution reveals a reasonably straight line. Several residuals seem to lie both below and above the line toward the respective ends of the plot indicating the presence of heavy negative and positive tails i.e. many large deviations from the predicted mean. However, the majority of the observations fall along the line and there isn't clear evidence of skewness. This intuition is confirmed by the plot of the empirical density of the model residuals. Again, we observe a somewhat large deviation from normality, but see that the empirical Student distribution captures the data somewhat more well.

To statistically assess the adequacy of the distribution we apply the augmented version of Pearson's test as discussed in Objective 1 with the null hypothesis of an adequately fitted Student's t distribution. The results (presented in Table 12) fail to reject an adequately fitted distribution at a 5% level for the each of the varying bin sizes. These results taken in conjunction with the visual evidence from the plots suggests that the Student's distribution describes the data well.

Table 12: Goodness of Fit Test for model residuals (d.f.=n-1)

Bin Group (n)	Statistic	p-value
20	26.94	.1061
30	29.92	.4179
40	50.51	.1026
50	60.38	.1277

Table 11 displays the results of Li-Mak's[29] ARCH LM test to determine the suitability of the GARCH fit.

Table 13: Li-Mak ARCH LM Test

Lag Length(m)	Test Statistic	d.o.f	p-value
2	1.344	1	.2462
3	1.404	2	.4956
4	1.957	3	.5812
5	1.973	4	.7408

Recall that the null hypothesis is of a well fit ARCH model and that the test statistic is asymptotically distributed chi-squared with *m-1* degrees of freedom. The test results fail to reject the null hypothesis of an adequately fit ARCH process for successively longer lag lengths implying that the model captures the behavior of process well.

Having determined that the model is well specified, we turn to interpretation and discussion. The first question we consider is the efficiency of the Bitcoin market. The consensus is that most financial markets exhibit semi-strong form efficiency. Semi-strong efficiency implies that all historical and publicly available information about an asset is reflected in the current market price. As such, in very weakly efficient market we would not expect a trader to be able to profit in the long run simply by knowing last week's return information. We can gauge this by examining the significance of the lagged return variable. This variable is not close to being significant which therefore confirms/reaffirms our conclusion made in Objective 1 that the Bitcoin market is weak form efficient.

With respect to the demand and supply side variables, we observe that the impact of changes in trading and transaction volume are no longer significant. This is a somewhat curious finding as we would think that weeks of heavy trading or transaction volumes would tend to move the price of Bitcoin more substantially than during low volume weeks, but that is not suggested by the data here. The hash-rate and the number of registered addresses continue to be highly significant variables. This supports the findings from our earlier model using daily data and confirms what we would expect from these 2 variables: that Bitcoin's price is determined in no small part by the number of users and the sustainability of the network.

Turning to the macro-financial variables we see that all proposed variables are insignificant at conventional levels. Taking each in turn, we observe that the current level of the Fed Funds rate does not significantly impact Bitcoin's price which conforms with our notion *a priori*. Bitcoin is not issued by a central authority (that's kind of the point) and does not exist alongside an interest rate regime. It follows that changes in Bitcoin's price would be unimpacted by the policies of an unrelated governing body. Insignificant coefficient estimates for changes in the Dollar-Yen and Dollar-Euro exchange rates suggest

Bitcoin is isolated from traditional currency markets. Because it is not issued by a government, holding Bitcoin is not a requisite for trade with a nation like the USD, Yen or Euro might be. Furthermore, it is not considered a reserve currency or reserve asset (like gold or other precious metals). This lack of relationship with major currencies implies that Bitcoin may not primarily function as a currency. This is a curious result indeed as there is known transaction volume and documented firms that accept BTC as a valid form of payment. Furthermore, since its inception, BTC has been widely adopted by black market operators due to its relative anonymity. The conclusion to draw from these observations seems to be that using Bitcoin as a currency or medium of exchange is possible, but not a primary driver of it's price determination.

Finally, for financial assets, a familiar way to assess market exposure or market risk is to examine how an asset's value fluctuates against a broad index like the S&P 500. The lack of a significant relationship between Bitcoin and the S&P 500 is quite striking and suggests that Bitcoin's return is not derived from a premium earned for taking traditional capital market risk. Taken in conjunction with the result for currency exposure it would seem the BTC is neither asset nor currency and rather acts as a source of absolute return or speculation which is notoriously difficult to assess but seems to support its' reputation as a speculative bubble.

Section C: Concluding Remarks on Objective 2

The focus of Objective 2 was to measure Bitcoin's exposure to macroeconomic and financial variables while accounting for supply and demand relationships believed to play an important role in Bitcoin's development. Using the variables of the S&P 500, USD-Yen exchange rate, USD-Euro exchange rate and the Federal Funds rate we developed a GARCH model to measure Bitcoin's response to changes in these factors. Model results indicate that Bitcoin is not meaningfully impacted by macro-financial factors and that its return process is better described by variables unique to the Bitcoin ecosystem like the hash rate and the number of users.

These results are consistent with those of Ciaian et al (2015)[12] who posited that Bitcoin is unimpacted by broad macro-financial variables once supply and demand variables are properly introduced. However, it runs counter to the findings of Van Wijk[11] and Dyhrberg[14] who suggested that macro variables are important when examining BTC's price return process but did not control for supply and demand effects.

CHAPTER 7: CONCLUSION

In this paper we provide an assessment of the market efficiency and financial time series characteristics of Bitcoin and accurately model the characteristic volatility using an GARCH specification. Furthermore, we address Bitcoin’s exposure to traditional macro-financial and asset specific variables which could plausibly be driving the return generating process. The analysis was conducted using both daily and weekly return data for Bitcoin spanning the period of September 14, 2011 to March 28, 2018.

Results from Objective 1 show Bitcoin is exposed to variables unique to its own ecosystem, namely: exchange volume, transaction volume, the number of unique addresses and the hash rate. We argued that the Bitcoin market is weakly efficient, and that past historical return data would not generate excess returns over a simple buy and hold investing strategy. Furthermore, we were able use these results to accurately back test and forecast Bitcoin’s value at risk and expected shortfall.

Results from Objective 2 provided further evidence in support of weak form market efficiency and the importance of asset specific variables (i.e. addresses and hash rate) in Bitcoin’s price development. We also demonstrated the Bitcoin is not significantly exposed to traditional factors of risk including: interest rates, market risk, or currency risk.

Taken together we conclude that while the Bitcoin market is efficient it essentially exists in parallel to conventional markets. This could be a result of continued relative obscurity and lack of relevant use cases that would drive further adoption among mainstream users. The results appear to confirm previous research that Bitcoin is viewed and traded primarily as a speculative asset rather a traditional currency or security. This finding suggests that there is a limited role for Bitcoin in an investor’s portfolio as it does not provide diversification or a hedge away from traditional risk factors and instead imbue a portfolio with additional idiosyncratic risk that would be difficult to diversify away.

REFERENCES

1. Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." (2008).
2. *CoinMarketCap Family of Sites*. CoinMarketCap 2018. www.coinmarketcap.com. Accessed 22 December 2018.
3. *"2017 Factbook."* Securities Industry and Financial Markets Association. SIFMA Research Department. www.sifma.org/research. Accessed 8 May 2018.
4. Thompson, Patrick. "Bitcoin Adoption by Businesses in 2017." *Coin Telegraph*, https://cointelegraph.com/news/bitcoin-adoption-by-businesses-in-2017. Accessed 8 May 2018.
5. Glaser, Florian, et al. "Bitcoin-asset or currency? revealing users' hidden intentions." *Revealing Users' Hidden Intentions (April 15, 2014). ECIS* (2014)
6. Yermack, David. "Is Bitcoin a real currency? An economic appraisal." *Handbook of digital currency*. Academic Press, 2015. 31-43.
7. Hayes, Adam S. "Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin." *Telematics and Informatics* 34.7 (2017): 1308-1321.
8. Chu, Jeffrey, Saralees Nadarajah, and Stephen Chan. "Statistical analysis of the exchange rate of bitcoin." *PloS one*10.7 (2015): e0133678.
9. Brown, Stephen J., and Jerold B. Warner. "Using daily stock returns: The case of event studies." *Journal of financial economics* 14.1 (1985): 3-31.
10. Urquhart, Andrew. "The inefficiency of Bitcoin." *Economics Letters* 148 (2016): 80-82.
11. Nadarajah, Saralees, and Jeffrey Chu. "On the inefficiency of Bitcoin." *Economics Letters* 150 (2017): 6-9.
12. Feng, Wenjun, Yiming Wang, and Zhengjun Zhang. "Informed trading in the Bitcoin market." *Finance Research Letters* 26 (2018): 63-70.
13. Bariviera, Aurelio F., et al. "Some stylized facts of the Bitcoin market." *Physica A: Statistical Mechanics and its Applications* 484 (2017): 82-90
14. Tiwari, Aviral Kumar, et al. "Informational efficiency of Bitcoin—An extension." *Economics Letters* 163 (2018): 106-109.
15. Van Wijk, Denis. "What Can Be Expected From Bitcoin." MA Thesis. *Erasmus Universiteit Rotterdam*. Published 2013.
16. Kristoufek, Ladislav. "BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era." *Scientific reports* 3 (2013): 3415
17. Ciaian, Pavel, Miroslava Rajcaniova, and d'Artis Kancs. "The economics of BitCoin price formation." *Applied Economics* 48.19 (2016): 1799-1815.

18. Barro, Robert J. "Money and the price level under the gold standard." *The Economic Journal* 89.353 (1979): 13-33.
19. Dyhrberg, Anne H. "Bitcoin, gold and the dollar–A GARCH volatility analysis." *Finance Research Letters* 16 (2016): 85-92.
20. Nelson, Daniel B. "Conditional heteroskedasticity in asset returns: A new approach." *Econometrica: Journal of the Econometric Society* (1991): 347-370.
21. Bouri, Elie, et al. "On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?" *Finance Research Letters* 20 (2017): 192-198.
22. Engle, Robert. "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models." *Journal of Business & Economic Statistics* 20.3 (2002): 339-350.
23. Matthew Hougan, Hong Kim, and Micah Lerner "Economic and Non-Economic Trading in Bitcoin: Exploring the Real Spot Market For The World's First Digital Commodity." *Bitwise Asset Management*. (2019).
24. Quandl. (2018). *Alternative Data* [database]. Retrieved from https://www.quandl.com/alternative-data
25. Yahoo! (2018). *Yahoo! Finance* [database]. Retrieved from https://finance.yahoo.com/
26. Federal Reserve Bank of St. Louis. (2018). *FRED® Economic Data* [database]. Retrieved from https://fred.stlouisfed.org
27. Ghalanos, Alexios. (2018). *rugarch*. Version 1.3-8. Comprehensive R Archive Network.
28. Engle, Robert F. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." *Econometrica: Journal of the Econometric Society* (1982): 987-1007
29. Bollerslev, Tim. "Generalized autoregressive conditional heteroskedasticity." *Journal of econometrics* 31.3 (1986): 307-327.
30. Dickey, David A., and Wayne A. Fuller. "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American statistical association* 74.366a (1979): 427-431.
31. Phillips, Peter CB, and Pierre Perron. "Testing for a unit root in time series regression." *Biometrika* 75.2 (1988): 335-346.
32. Mandelbrot, Benoit. "The Variation of Certain Speculative Asset Prices." *The Journal of Business* 36.4 (1963): 394-419.
33. Li, W. K., and T. K. Mak. "On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity." *Journal of Time Series Analysis* 15.6 (1994): 627-636.
34. Box, George EP, and David A. Pierce. "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models." *Journal of the American statistical Association* 65.332 (1970): 1509-1526.

35. Ljung, Greta M., and George EP Box. "On a measure of lack of fit in time series models." *Biometrika* 65.2 (1978): 297-303.
36. Palm, Franz C. "7 GARCH models of volatility." *Handbook of statistics* 14 (1996): 209-240.
37. Engle, Robert F., and Victor K. Ng. "Measuring and testing the impact of news on volatility." *The Journal of Finance* 48.5 (1993): 1749-1778.
38. Kupiec, Paul. "Techniques for verifying the accuracy of risk measurement models." *The J. of Derivatives* 3.2 (1995).
39. Christoffersen, Peter, and Denis Pelletier. "Backtesting value-at-risk: A duration-based approach." *Journal of Financial Econometrics* 2.1 (2004): 84-108.

APPENDIX

In this section we briefly extend our investigation of weak form efficiency for the Bitcoin market. As mentioned previously, weak form market efficiency implies that future returns cannot be effectively predicted based on past information. In the above analysis for Objectives 1 and 2 we leaned on the literature and assumed weak form efficiency. As such the impacts of supply-demand and macro-financial variables were analyzed contemporaneously; i.e. the impact of a change in the S&P 500 index at time *t* is reflected in the change of the market price of Bitcoin also at time *t.* However, in case that the BTC market is not weak form efficient, then lagged values of the independent variables could be meaningful and help to explain Bitcoin returns in the present.

The analysis to follow was conducted for both daily return data and weekly return data; i.e. the data used in the investigation of Objective 1 and Objective 2 respectively. Each analysis consists of 3 models with varying lag lengths of the independent variables introduced previously. The model specifications are described in Table A1.

Table A1: Model and Lag Specification

Model	Model Abbreviation	Number of Return Lags	Number of Non-return Lags
GARCH 1	V1	1	1
GARCH 2	V2	2	1
GARCH 3	V3	2	2

The results for each model using daily data (corresponding to Objective 1's variable set) are presented in Table A2. In the table, the standard error is found is parentheses under the estimate followed by the p-value of the test statistic.

Table A2: Estimation Results for daily GARCH (1,1) with varying lag lengths

Variable	V1	V2	V3
Constant	-.0659 (.0297) p=.0270	-.0689 (.0299) p=.021	-.0682 (.0304) p=.0252
AR (1)	-.0195 (.0205) p=.3416	-.0213 (.0205) p=.2975	-.0211 (.0205) p=.3021
AR (2)		-.0386 (.0206) p=.0604	-.0374 (.0205) p=.0692
ExVol (1)	-.0009 (.0017) p=.5897	-.00102 (.0017) p=.5650	-.0004 (.0019) p=.8293
ExVol (2)			.0012 (.0019) p=.5163
TransVol (1)	.0011 (.0078) p=.891	.0018 (.0031) p=.8224	-.0004 (.0099) p=.9640
TransVol (2)			.0018 (.0087) p=.8212
Addresses (1)	.0075 (.0031) p=.0181	.0078 (.0031) p=.0135	.0065 (.0099) p=.5145
Addresses (2)			.0012 (.0099) p=.8978
Hash Rate (1)	-.0019 (.0007) p=.0142	-.0021 (.0008) p=.0101	.0097 (.0085) p=.2577
Hash Rate (2)			-.0117 (.0085) p=.1696
Omega	.00003 (.000016) p=.0157	.00004 (.000016) p=.0172	.00003 (.00001) p=.01636
Alpha	.1986 (.0291) p<.001	.1974 (.0293) p<.001	.1948 (.0286) p<.001
Beta	.8003 (.0311) p<.001	.8015 (.0341) p<.001	.8041 (.0305) p<.001
Shape	.8181 (.0342) p<.001	.815 (.0339) p<.001	.8133 (.034) p<.001

	V1	V2	V3
Log-Likelihood	4177.137	4178.663	4180.908
Bayes-Schwartz IC	-4.1129	-4.1164	-4.1142

Examining the results, a couple of features immediately stand out. To begin, the estimated coefficient for the AR (1) variable (i.e. the 1-day lag return) is not statistically significant for each of the 3 models. However, models V2 and V3 indicate the AR (2) term is statistically significant (albeit, only at a 10% level of confidence). This contrasts with the results we observed during the examination of Objective 1 where it was observed that AR (1) and AR (2) were highly statistically significant. However, in Objective 1 we argued weak efficiency by employing Fama's submartingale argument stating that despite the significance of AR (1) and AR (2) they account for so little predictive power that a trading rule based solely on historical data would be unlikely to outperform a simple buy and hold strategy. That argument appears to be strengthened here as we observe that if lags of the other model variables are included, the predictive power of the lagged returns is further eroded.

For the supply and demand variables we observe that across the three models the lagged exchange and trading volume variables are not statistically significant. This result makes intuitive sense as we would not anticipate the volume from yesterday to have a significant impact on the price of Bitcoin today; except possibly under very stressed market conditions where there is an unusual amount of buying or selling. Recall that both exchange volume and transaction volume when measured at time 't' were significant in Objective 1.

The results for the number of registered addresses and the hash rate offer the most interesting results. In models V1 and V2 we observe that both the lagged number of addresses and lagged hash rate are significant. These results align with those observed in Objective 1 and confirm the importance of these variables to Bitcoin's development. However, when additional lags are included in model V3 we see that Addresses (1) and Hash Rate (1) are no longer significant; nor are their counterparts at lag length 2. It is possible that the lagged values of Addresses and Hash Rate generally offer some predictive power, but when multiple lags are included the effect is more difficult to detect and ascribe.

To evaluate which of the three models is best we perform a series of likelihood ratio tests. The likelihood ratio is of the form:

$$LR = -2 * \ln(\frac{\theta_R}{\theta_{UR}})$$

Where θ_{UR} and θ_R denote the likelihood of unrestricted and restricted models respectively. The null hypothesis of test is that the restricted model (i.e. the simpler model) is the correct model. The LR test statistic has been shown to be asymptotically distributed χ_m^2 where m is the difference between the number of estimated parameters in the two models. To test which model is best we will proceed by moving from most complex to most general. Model V3 (being the most complex) will first be tested against model V2. The results of the testing procedure are shown in Table A3.

Table A3: LR Test for V2 and V3

Model	Type	Bayes IC	Log-Likelihood	LR	d.f.	P(X>LR)	Decision
V2	Restricted	-4.1164	4178.663	4.49	4	.3437	Fail to reject
V3	Unrestricted	-4.1142	4180.908				

The test results imply that model V2 is superior to V3. We now proceed to determine if the additional return lag in V2 is superior to V1. The results are presented in Table A4.

Table A4: LR Test for V1 and V2

Model	Type	Bayes IC	Log-Likelihood	LR	d.f.	P(X>LR)	Decision
V1	Restricted	-4.1129	4177.137	3.052	1	.0806	Reject
V2	Unrestricted	-4.1164	4178.663				

The results of the test are weakly significant (at a 10% confidence level) and suggest that model V2 is slightly better than model V1. Furthermore, because model V2 minimizes the Bayes IC we feel comfortable with including the additional lagged return variable and conclude that model V2 is the best fit model among our candidate models. This is an encouraging result as the V2 specification aligns most closely with that of Objective 1.

We now proceed to conduct a similar analysis for the weekly data corresponding to Objective 2. Recall that the focus of Objective 2 was to assess the impact that macro-financial variables have on Bitcoin. Again, we present 3 candidate models with specifications corresponding to those detailed in Table A1. The results from the estimation are presented in Table A5.

Table A5: Estimation Results for weekly GARCH (1,1) with varying lag lengths.

Variable	V1	V2	V3
Constant	-.4334 (.2534) p=.0882	-.3567 (.2624) p=.1750	-.3827 (.2641) p=.1483
AR (1)	.0212 (.0576) p=.7128	.0227 (.0582) p=.6966	.0202 (.0598) p=.7352
AR (2)		.0862 (.0569) p=.1308	.0565 (.0576) p=.3272
ExVol (1)	.0121 (.0094) p=.2004	.0107 (.0094) p=.2535	.0218 (.0101) p=.0336
ExVol (2)			.0218 (.0102) p=.0336
TransVol (1)	-.0752 (.0424) p=.0772	-.0722 (.0426) p=.0914	-.0015 (.0541) p=.9785
TransVol (2)			.0725 (.0449) p=.1083
Addresses (1)	.0512 (.0271) p=.0597	.0425 (.0281) p=.1304	-.0147 (.0701) p=.8342
Addresses (2)			.0595 (.0693) p=.3910
Hash Rate (1)	-.0156 (.0068) p=.0223	-.0133 (.0071) p=.0610	-.0274 (.0514) p=.9785
Hash Rate (2)			.0137 (.0514) p=.7898
Fed Funds (1)	.0313 (.0228) p=.1705	.0281 (.0229) p=.2211	-.4117 (.1625) p=.0118

Fed Funds (2)			.4478 (.1642) p=.0068
Change JPY (1)	-.5939 (.5894) p=.3144	-.6049 (.5899) p=.3060	-.6405 (.5872) p=.2763
Change JPY (2)			.2553 (.5942) p=.6677
Change EUR (1)	-1.081 (.6444) p=.0941	-1.041 (.6482) p=.1091	-1.4039 (.6526) p=.0322
Change EUR (2)			.3841 (.6429) p=.5507
SPX (1)	.4541 (.4746) p=.3395	.4135 (.4763) p=.3860	.3936 (.4786) p=.4115
SPX (2)			.7209 (.4784) p=.1328
Omega	.0011 (.0006) p=.1005	.0011 (.0007) p=.1190	.0028 (.0015) p=.0683
Alpha	.3481 (.1014) p<.001	.3451 (.1103) p=.0017	.2437 (.0863) p=.0047
Beta	.6099 (.0767) p<.001	.6132 (.0785) p<.001	.5583 (.1509) p<.001
Shape	1.2543 (.01402) p<.001	1.1644 (.1539) p<.001	1.2618 (.1315) p<.001

	V1	V2	V3
Log-Likelihood	268.5479	269.1243	270.2729
Bayes-Schwartz IC	-1.7067	-1.7164	-1.5423

The results from Models V1 and V2 largely coincide. Neither model shows the return lags to a have a significant effect on Bitcoin price. Interestingly, Model V1 does show the lagged Addresses and Hash Rate variables have a significant impact on BTC. These results are broadly consistent with the estimates obtained in Objective 2 and, again, suggest the continuing importance of these variables to the BTC ecosystem. The power of these results is eroded to a degree when an additional return lag is included in model V2 as we see lagged Addresses is no longer significant and the Hash Rate is more weakly significant. Neither model shows a substantive relationship between Bitcoin returns and the lagged macro-financial variables; which seems to confirm the intuition gleaned from Objective 1.

However, the results from Model V3 contrast with those of V1 and V2. V3 suggests that no significant relationship exists between Bitcoin and supply-demand variables, but, curiously, a significant relationship is observed for both lags of the Fed Funds rate. This seems to contradict the results of both Objective 2 and the two sister models V1 and V2. Additionally, the impact of the change in the EUR-USD exchange rate seems to become meaningful.

Based on the above results, an assessment of the relative informational efficiency of the models seems prudent to determine which is best. Tables A6 and A7 contain the results for the likelihood ratio tests. Again, we proceed from most to least complex.

Table A6: LR Test for models V2 and V3

<table>
<tr><th>Model</th><th>Type</th><th>Bayes IC</th><th>Log-Likelihood</th><th>LR</th><th>d.f.</th><th>P(X>LR)</th><th>Decision</th></tr>
<tr><td>V2</td><td>Restricted</td><td>-1.7164</td><td>269.1243</td><td rowspan="2">2.29</td><td rowspan="2">8</td><td rowspan="2">.97</td><td rowspan="2">Fail to reject</td></tr>
<tr><td>V3</td><td>Unrestricted</td><td>-1.5423</td><td>270.2729</td></tr>
</table>

The test results imply that model V2 is superior to V3. We now proceed to determine if the additional return lag in V2 is superior to V1. The results are presented in Table A7.

Table A7: LR Test for models V1 and V2

<table>
<tr><th>Model</th><th>Type</th><th>Bayes IC</th><th>Log-Likelihood</th><th>LR</th><th>d.f.</th><th>P(X>LR)</th><th>Decision</th></tr>
<tr><td>V1</td><td>Restricted</td><td>-1.7067</td><td>268.5479</td><td rowspan="2">1.15</td><td rowspan="2">1</td><td rowspan="2">.28</td><td rowspan="2">Fail to reject</td></tr>
<tr><td>V2</td><td>Unrestricted</td><td>-1.7164</td><td>269.1243</td></tr>
</table>

The test results suggest that the AR (2) variable present in model V2 does not add significant predictive power over model V1. Hence, we conclude that V1 is the best model among our choice of candidate models for weekly data.