# Understanding and Modeling the Factors Influencing the Price of Bitcoin



By David Sheffet

EC 431

**Table of Contents**

## Abstract

The first decentralized digital currency was called Bitcoin. Over the past ten years, its value and popularity have grown significantly; analysts, traders, and researchers are all very interested in projecting its price. I thoroughly review models for predicting the price of Bitcoin in this paper. We investigate the efficacy of various pre-processing features and methods, such as technical indicators and social media sentiment analysis. I use a variety of metrics, including Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), to assess the effectiveness of the models. We also provide insights into the most significant features that influence Bitcoin price and discuss this research's limitations and future directions. Overall, this study contributes to the ongoing research in Bitcoin price prediction and provides practical implications for investors and traders.

## JEL Codes

C53: Forecasting and Prediction Methods

G17: Financial Forecasting and Simulation

C58: Financial Econometrics

## Introduction

Bitcoin is a digital currency, also called a virtual currency, which can be transacted for a low cost nearly instantly from anywhere in the world. Bitcoin also powers the blockchain, a public, immutable, and decentralized global ledger. Unlike traditional currencies such as dollars, bitcoins are issued and managed without the need for any central authority whatsoever. There is no government, company, or bank in charge of Bitcoin. As such, it is more resistant to wild inflation and corrupt banks. With Bitcoin, you can be your own bank. Read the Bitcoin whitepaper to understand further the schematics of how Bitcoin works.[1]

Thus, giving Bitcoin unique status as a financial commodity; It has unique price mechanisms. Bitcoin's volatility has been a topic of debate since its inception. The cryptocurrency's value is subject to frequent fluctuations due to various factors such as market demand, regulatory changes, and investor sentiment. However, recent data shows that Bitcoin's volatility has decreased significantly compared to its early years. Some analysts suggest that Bitcoin is now less volatile than certain well-established stocks, such as Amazon. This suggests that Bitcoin may be gradually maturing as an asset class and could become a more attractive investment option for risk-averse investors.

This begs the question, what drives Bitcoin's price?

## Literary Review

The basic principle behind Bitcoin is supply and demand; however, there is significant uncertainty regarding the price. One of the first papers to shed some light on this was the economics of bitcoin price formation.

2010 through 2015 are covered in the article published in 2016 by The Journal Applied Economics. In their 2016 paper "The Economics of BitCoin Price Formation," authors Pavel Ciaian, Miroslava Rajcaniova, and d'Artis Kancs look at the economics of Bitcoin price growth. The authors examine how Bitcoin's market structure differs from

[1] https://bitcoin.org/bitcoin.pdf

established financial markets, which state institutions and laws heavily influence. They claim that the value of Bitcoin comes from its usefulness as a medium of trade and a form of wealth storage and that the relationship between its supply and demand determines its price. The authors use econometric methods to investigate how market sentiment, technological improvements, and macroeconomic factors affect Bitcoin's price.

The paper begins by giving a general overview of the Bitcoin market and describing how it differs from other marketplaces. According to the creators, Bitcoin is a decentralized digital currency that any government or financial organization does not support. This indicates that conventional economic variables like inflation or interest rates do not impact pricing.

The writers then look at the factors that affect Bitcoin's price, such as demand and supply, market sentiment, and adoption. They use econometrics to show that the supply of Bitcoin has a considerable impact on its price, with a 1% increase in supply leading to a 0.56% decrease in price. The researchers also find that the price of Bitcoin is significantly influenced by market sentiment, which is determined by Google searches and social media mentions. For instance, a 1% increase in Bitcoin searches on Google causes a 0.25% price increase.

The authors also look at how adoption affects the development of the price of Bitcoin. They discover that the quantity of Bitcoin wallets, the quantity of Bitcoin transactions, and the quantity of Bitcoin-accepting businesses are all connected favorably with the price of Bitcoin. For instance, a 1% increase in Bitcoin wallets causes a 0.39% price increase.

One interesting finding of the article is that the price of Bitcoin is affected by external events such as regulatory announcements and security breaches. For instance, the authors discover that after Mt. Gox, a prominent Bitcoin exchange, lost 850,000 Bitcoins as a result of a security breach and filed for bankruptcy, the price of Bitcoin dropped considerably. The authors also discovered that the price of Bitcoin was significantly impacted negatively by regulatory statements, such as the Chinese government's announcement that it would forbid financial institutions from trading in Bitcoin.

Additional resources that have been extremely helpful are E*TRADE's public materials on the factors that affect Bitcoin. In particular, its discussion of how speculation and people's tolerance to risk significantly determines the price.[2]

Kade Garrett's paper "What determines the price of Bitcoin" was also beneficial. She specifically talked about how inflation and foreign exchange rates, in the feds, can directly influence the price of crypto. Additionally, she talked about how centralized crypto exchanges play a significant role in price discovery.[3]

That said, there is a clear comprehension and rigorous analysis of the economics of Bitcoin price formation, which is helpful for researchers, investors, and policymakers interested in cryptocurrency, particularly by Pavel Ciaian. They provide a clear and concise overview of Bitcoin's market structure, the factors that influence its price, and a detailed description of their econometric methodology. These authors' results highlight the importance of market sentiment and network effects in explaining Bitcoin's price dynamics, which is valuable to the literature.

---

[2] https://us.etrade.com/knowledge/library/cryptocurrency/factors-that-affect-bitcoin-price

[3] https://decrypt.co/resources/what-determines-the-price-of-bitcoin-2

That being said, the more in-depth data is often based on data from a relatively short period, which limits the generalizability of their results. Additionally, their study focuses only on Bitcoin and does not explore the dynamics of other cryptocurrencies. This is an inherent problem with Bitcoin, as the technology is still in its infancy. However, it is a problem in the less.

For instance, Pavel Ciaian only focuses on Bitcoins price thru 2016, and a lot has changed since then. Finally, other than E-Trade their analysis does not explicitly consider the potential impact of government regulations on Bitcoin's price, which is a significant factor in traditional financial markets.

While someone argues that the government cannot control Bitcoin, and then, even if they wanted to, they would not be able to kill Bitcoin, they definitely can influence the price and, therefore, much to the chagrin of many crypto fans still need to be involved in the analysis of potential price shocks.

They all collectively also seem to underestimate and undervalue the effects of significant events involving Bitcoin, such as FTX's collapse, due to the inherent unpredictability of these events. Finally, their analysis does not explicitly consider the potential impact of government regulations on Bitcoin's price, which is a significant factor in traditional financial markets.

Ciaian, Rajcaniova, and Kancs' article is the most useful for this project and is a valuable contribution to the literature on Bitcoin and cryptocurrency more broadly. The authors offer a clear and rigorous analysis of Bitcoin's price formation, and their results provide insights into the drivers of the cryptocurrency's value. However, their study has some limitations, particularly regarding the generalizability of their results and the need for more consideration of government regulations on Bitcoin's price.

## Economic Theory

With Bitcoin, there is no need for middlemen like banks or payment processors to facilitate transactions between individuals or corporations. Concerning cross-border transactions, in particular, this capability makes Bitcoin transactions more affordable, quick, and secure than conventional payment methods. Contrary to fiat currencies, the 21 million unit restriction on the supply of bitcoin ensures that its value will not be diminished by inflation.

Bitcoin and gold are frequently contrasted as a store of value since both have finite supplies and are independent of any government or central bank. Like gold, bitcoin's value is mainly determined by its rarity and the demand from investors looking to protect themselves from inflation or geopolitical concerns. However, bitcoin is much more flexible and available as a store of value than gold because it is easily divisible and can be transmitted and received instantly.

Since most prices and contracts are still in fiat currency, bitcoin is still in its infancy as a unit of account. However, some companies and individuals are beginning to use Bitcoin to account for the cost of products and services, particularly in sectors like e-commerce and online gaming, due in part to Bitcoin's low transaction fees and fast settlement.

Thus initially, I anticipated that the equation to price Bitcoin would be something along the lines of:

Predicted Equation:

Price = β0 + β1*MarketCap + β2*Volume + β3*Blocksize + β4*MiningDiff + β5*HashRate + β6*SocialMedia + β7*WalletCreation + ε

Variables

Market Capitalization: This variable is the total market value of all the Bitcoin in circulation. The type of variable is quantitative and continuous. We can obtain the data from various sources such as coinmarketcap.com, blockchain.info, or Yahoo Finance.

Trading Volume: This variable is the total number of Bitcoins traded on various exchanges over a certain period. The type of variable is quantitative and continuous. We can obtain the data from various sources such as coinmarketcap.com, blockchain.info, or Yahoo Finance.

I use daily trading volume and the total number of bitcoins on any given day. This allows me to factor in the number of bitcoins left to mine and the number of bitcoins mined daily. This does not include the number of bitcoins that have been lost; however, that number is unknown.

Additionally, block details are essential to factor in. So there is a used average block size, the number of transactions per block, and the current block height. The estimated time to have it could be considered a weakness; however, while predictable, it has yet to be precisely known when that will be, although that will have a marked effect on the price.

Mining difficulty: the variable that measures how hard it is to mind a new block of Bitcoin. For this, I used the heart rate and the rate of difficulty. We can obtain the data from various sources such as blockchain.info, Bitcoinwisdom.com, or Bitinfocharts.com.

Sentiment Analysis: This variable measures the overall sentiment towards Bitcoin on social media, news, or other online platforms. The type of variable is qualitative, ordinal, or binary. We can obtain data from various sources like Twitter, Reddit, or Google Trends.

Additionally, the Adoption Rate is a factor. This variable measures the adoption rate of Bitcoin by individuals and businesses. The type of variable is quantitative, continuous, or categorical. We can obtain the data from various sources such as surveys, blockchain.info, or Bitcoincharts.com.

## Data

The primary place I got data from was blockchain.com. I obtained the time information from blockchain.com and the pricing information from the crypto exchange Gemini. I obtain the market information from Gemini. I obtain the trading volume information—specifically, the total daily trading volume and the number of bitcoins from blockchain.com. I am changing the black detail information from Blockchain.com, including the average block size and the number of transactions per block. I obtained information on mining difficulty from blockchain.com, including the cash rate and difficulty adjustment information. I am changing my sentiment analysis information from Google Trends. I obtained user adoption information from blockchain.com. I additionally got economic indicator data from FRED[4].

Some statistics had to be calculated using the data as it appeared in its original form to transform the data so it would be in a form that could be compared across seasons, including the market capitalization data.

[4] fred.stlouisfed.org

The data was from a mixture of Excel, compatible, CSV, and JSON files. The data was incompatible and had to be transformed to fit into its current format. An example of this should be market price, which was only taken every three days, as it was an average of the exchange prices, which constantly fluctuate. Trade volume had to be transformed to fit its current format. Additionally, things like hash rate block, size, etc., can only be done every three days, given that I went off the Gemini pricing data. An in-depth equation is required for the market control score as it measures mining and poor power diversity. When crypto is mined, it is more profitable to do this in groups as it makes it more likely that you all find bitcoin, mining pools control, and a Norton amount of power in terms of the mining network; however, it is essential that no mining pool becomes significantly more powerful than the others and obtains significant control over the network. Where is the control score a measure of the average amount of control that the top five mining pools have?

The web search percentile also had to be edited as Google Trends only provides data monthly. Thus I had to assume, most likely falsely, that the monthly average is constant throughout the month.

Additionally, I needed help combining valuable data, such as the amount of money on the exchanges. This is important in discussing exchange liquidity, which is significant in Bitcoin pricing. The number of bitcoins could also not be connected to the primary data set. This is because the number of Bitcoins mined, or on a given exchange, at any moment is only finalized in accuracy after mining a block. Thus these two metrics are on block time instead of minutes or days. Unfortunately, there was no overlap between when a block was in mind and when the day ended. Therefore, they are on two separate sheets inside the same Excel file. In the following analysis, they are you separately to conduct two different methods of pricing Bitcoin.

For more information on the variables, view the appendix below.

## Results

### Data Set 1

Table 1:

```
. regress MrktPrice HashRate NTransactionsperblock Difficulty NVT NWalletperd NetworkControlScore TradeVol VolumeBTC VolumeUSD WebSear
> chesPrecentile ntransactionsexcludingpopular open
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 1.2292e+09 | 12 | 102435466 |
| Residual | 139685.662 | 84 | 1662.92455 |
| Total | 1.2294e+09 | 96 | 12805888.4 |

| | | |
|---|---|---|
| Number of obs | = | 97 |
| F(12, 84) | = | 61599.59 |
| Prob > F | = | 0.0000 |
| R-squared | = | 0.9999 |
| Adj R-squared | = | 0.9999 |
| Root MSE | = | 40.779 |

| MrktPrice | Coefficient | Std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| HashRate | -8.85e-07 | 1.13e-06 | -0.79 | 0.434 | -3.12e-06 | 1.35e-06 |
| NTransactionsperblock | -.0812223 | .1513513 | -0.54 | 0.593 | -.382201 | .2197564 |
| Difficulty | 2.23e-12 | 8.62e-12 | 0.26 | 0.797 | -1.49e-11 | 1.94e-11 |
| NVT | -.3484483 | 1.097325 | -0.32 | 0.752 | -2.5306 | 1.833703 |
| NWalletperd | .0002416 | .0002093 | 1.15 | 0.252 | -.0001747 | .0006579 |
| NetworkControlScore | .800033 | 1.600642 | 0.50 | 0.619 | -2.383018 | 3.983085 |
| TradeVol | -5.29e-08 | 4.80e-08 | -1.10 | 0.274 | -1.48e-07 | 4.26e-08 |
| VolumeBTC | .1052328 | .0357236 | 2.95 | 0.004 | .0341925 | .176273 |
| VolumeUSD | -6.48e-06 | 2.00e-06 | -3.23 | 0.002 | -.0000105 | -2.49e-06 |
| WebSearchesPrecentile | .6309367 | 2.495559 | 0.25 | 0.801 | -4.331757 | 5.59363 |
| ntransactionsexcludingpopular | .0006411 | .0010252 | 0.63 | 0.533 | -.0013975 | .0026798 |
| open | 1.005125 | .0022342 | 449.88 | 0.000 | 1.000682 | 1.009568 |
| _cons | -120.6115 | 109.9709 | -1.10 | 0.276 | -339.3006 | 98.07773 |

I could assemble a regression analysis that fit the model exceptionally well using the first data set I assembled and the multiple linear regression model. The R-squared value of 0.9999 indicates that the model explains a large proportion of the variation in the bitcoin price. In this model, "open" is the most important predictor of "MrktPrice", with a coefficient of 1.005, indicating that a one-unit increase in "open" is associated with a $1.005 increase in "MrktPrice". "VolumeUSD" is also significant, with a negative coefficient of -6.48e-06, indicating that a one-unit increase in "VolumeUSD" is associated with a decrease in "MrktPrice" of 6.48e-06. Other variables, such as

"HashRate", "NTransactionsperblock", "Difficulty", "NVT", "TradeVol", "WebSearchesPrecentile", and "ntransactionsexcludingpopular" are not significant predictors of "MrktPrice" at the 5% significance level, as their p-values are more significant than 0.05.

This results in

Equation 1:

MrktPrice = -120.6115 + (-8.85e-07 * HashRate) + (-0.0812223 * NTransactionsperblock) + (2.23e-12 * Difficulty) + (-0.3484483 * NVT) + (0.0002416 * NWalletperd) + (0.800033 * NetworkControlScore) + (-5.29e-08 * TradeVol) + (0.1052328 * VolumeBTC) + (-6.48e-06 * VolumeUSD) + (0.6309367 * WebSearchesPrecentile) + (0.0006411 * ntransactionsexcludingpopular) + (1.005125 * open) + ε

While this model was very accurate, I wanted more than a model with many statistical and significant variables. Additionally, I did not like that the starting price was the leading and most important predictor. I wanted to find a methodology that allowed me to predict the price accurately without including previous price data.

Table 2:

```
. regress MrktPrice VolumeUSD VolumeBTC NWalletperd TradeVol HashRate NTransactionsperblock
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 860185601 | 6 | 143364267 |
| Residual | 369179683 | 90 | 4101996.47 |
| Total | 1.2294e+09 | 96 | 12805888.4 |

| | | |
|---|---|---|
| Number of obs | = | 97 |
| F(6, 90) | = | 34.95 |
| Prob > F | = | 0.0000 |
| R-squared | = | 0.6997 |
| Adj R-squared | = | 0.6797 |
| Root MSE | = | 2025.3 |

| MrktPrice | Coefficient | Std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| VolumeUSD | .0006209 | .0000646 | 9.62 | 0.000 | .0004927 | .0007492 |
| VolumeBTC | -11.28164 | 1.151493 | -9.80 | 0.000 | -13.56928 | -8.993995 |
| NWalletperd | -.0170661 | .0060905 | -2.80 | 0.006 | -.029166 | -.0049661 |
| TradeVol | 6.53e-06 | 2.15e-06 | 3.04 | 0.003 | 2.26e-06 | .0000108 |
| HashRate | .0000395 | 5.33e-06 | 7.41 | 0.000 | .0000289 | .0000501 |
| NTransactionsperblock | 4.193618 | .9436021 | 4.44 | 0.000 | 2.318988 | 6.068249 |
| _cons | 11839.41 | 2542.307 | 4.66 | 0.000 | 6788.677 | 16890.15 |

Using this same data set I was able to create a multiple linear regression model that fits the model well. In this model, the R-squared value is 0.6997, which suggests that the independent variables collectively explain about 70% of the variance in the market price of Bitcoin. The adjusted R-squared value is slightly lower at 0.6797, indicating that the addition of more independent variables may not significantly improve the model's fit. In this model all six independent variables, (VolumeUSD, VolumeBTC, NWalletperd, TradeVol, HashRate, and NTransactionsperblock) are statistically significant at a significance level of 0.05 or lower.

This results in

Equation 2:

MrktPrice = (0.0006209 * VolumeUSD) - (11.28164 * VolumeBTC) - (0.0170661 * NWalletperd) + (6.53e-06 * TradeVol) + (0.0000395 * HashRate) + (4.193618 * NTransactionsperblock) + 11839.41 + ε

In this equation, I wanted to make sure that I did not suffer from omitted variable bias, due to the fact that the variables included are all bitcoin related variables. That's why I decided an additional model was an order and felt the need to include some traditional financial metrics.

Table 3:

```
. regress MrktPrice VolumeUSD VolumeBTC NWalletperd TradeVol HashRate NTransactionsperblock VIX FFR
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 899544713 | 8 | 112443089 |
| Residual | 263552254 | 86 | 3064561.09 |
| Total | 1.1631e+09 | 94 | 12373372 |

| | |
|---|---|
| Number of obs | = 95 |
| F(8, 86) | = 36.69 |
| Prob > F | = 0.0000 |
| R-squared | = 0.7734 |
| Adj R-squared | = 0.7523 |
| Root MSE | = 1750.6 |

| MrktPrice | Coefficient | Std. err. | t | P>\|t\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| VolumeUSD | .0004025 | .0000745 | 5.40 | 0.000 | .0002543 | .0005507 |
| VolumeBTC | -7.531505 | 1.276853 | -5.90 | 0.000 | -10.0698 | -4.993206 |
| NWalletperd | -.0215044 | .0054132 | -3.97 | 0.000 | -.0322654 | -.0107433 |
| TradeVol | 6.86e-06 | 1.87e-06 | 3.66 | 0.000 | 3.13e-06 | .0000106 |
| HashRate | .000062 | 7.53e-06 | 8.24 | 0.000 | .000047 | .000077 |
| NTransactionsperblock | 5.865124 | 1.070299 | 5.48 | 0.000 | 3.737439 | 7.992808 |
| VIX | -271.2442 | 61.74447 | -4.39 | 0.000 | -393.9882 | -148.5003 |
| FFR | -2014.362 | 471.3613 | -4.27 | 0.000 | -2951.397 | -1077.327 |
| _cons | 19742.04 | 3108.402 | 6.35 | 0.000 | 13562.74 | 25921.34 |

This additional output includes a variable for the federal funds rate, and a variable for the VIX index. This gives us a much better idea of how governmental policy can affect the price of bitcoin and helps show just how much of a role they do in fact play.

The output shows that the model is statistically significant, as indicated by the low p-value. The R-squared value is 0.7734, which means that the independent variables in the model explain about 77.34% of the variance in the dependent variable. While all 8 variables are statistically significant "VolumeUSD," "VolumeBTC," "NWalletperd," "HashRate," "NTransactionsperblock," "VIX," and "FFR" have negative coefficients, while "TradeVol" has a positive coefficient. Resulting in the following equation.

Equation 3:

$$MrktPrice = \beta 0 + \beta 1 VolumeUSD + \beta 2 VolumeBTC + \beta 3 NWalletperd + \beta 4 TradeVol + \beta 5 HashRate + \beta 6 NTransactionsperblock + \beta 7 VIX + \beta 8 FFR + \varepsilon$$

or

$$MrktPrice = 19742.04 + 0.0004025 * VolumeUSD - 7.531505 * VolumeBTC - 0.0215044 * NWalletperd + 6.86e{-}06 * TradeVol + 0.000062 * HashRate + 5.865124 * NTransactionsperblock - 271.2442 * VIX - 2014.362 * FFR + \varepsilon$$

Where:

MrktPrice is the predicted market price of the cryptocurrency

VolumeUSD is the total USD volume of transactions in the past 24 hours

VolumeBTC is the total BTC volume of transactions in the past 24 hours

NWalletperd is the number of unique active addresses per day

TradeVol is the total number of trades in the past 24 hours

HashRate is the estimated number of tera hashes per second (trillions of hashes per second) the Bitcoin network is performing

NTransactionsperblock is the total number of transactions in each block

VIX is the CBOE Volatility Index, a measure of stock market volatility

FFR is the Federal Funds Rate, the interest rate at which depository institutions lend to each other overnight

e is the error term representing the variation in MrktPrice that is not explained by the independent variables.

Data Set 2

Table 2:

```
. regress Price Valueonexchanges ofBTCs

      Source |       SS           df       MS      Number of obs   =     1,500
-------------+----------------------------------   F(2, 1497)      >  99999.00
       Model |  2.9357e+11         2  1.4679e+11   Prob > F        =    0.0000
    Residual |   130609933     1,497  87247.7843   R-squared       =    0.9996
-------------+----------------------------------   Adj R-squared   =    0.9996
       Total |  2.9370e+11     1,499   195932058   Root MSE        =    295.38

------------------------------------------------------------------------------
           Price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----------------+----------------------------------------------------------------
Valueonexchanges |   5.28e-08   3.42e-11  1543.59   0.000     5.27e-08    5.28e-08
          ofBTCs |   .0000211   1.63e-06    12.98   0.000     .0000179    .0000243
           _cons |   -93.0753   20.78255    -4.48   0.000    -133.8413   -52.30928
------------------------------------------------------------------------------
```

The above table is a regression model for:

Equation 4:

$$Price = \beta 0 + \beta 1 * Valueonexchanges + \beta 2 * ofBTCs + \varepsilon$$

or

$$Price = -93.0753 + 5.28e\text{-}08(Valueonexchanges) + 0.0000211(ofBTCs) + \varepsilon$$

This regression result indicates that both independent variables are significant predictors of the dependent variable. In this instance, for every one unit increase in "Valueonexchanges," there is a predicted increase in "Price" of 5.28e-08, holding "ofBTCs" constant. For every one unit increase in "ofBTCs," there is a predicted increase in "Price" of 0.0000211, holding "Valueonexchanges" constant. This shows that supply and demand principles have determined the price of Bitcoins. At the same time, this model's variables are limited. It is fascinating as there was a lot more data as this is a daily model from every day since bitcoins inception. This provides a unique perspective on how vital supply and demand principles are to Bitcoin. This is the case since the model explains a significant amount of the variation in "Price," with an R-squared value of 0.9996. The root means square error (RMSE) is 295.38, which indicates the average difference between the observed and predicted values of "Price."

## Conclusion

Overall, the results of this model found that TradeVol is the most positively correlated, followed by NTransactionsperblock. VolumeUSD, VolumeBTC, NWalletperd, HashRate, NTransactionsperblock, VIX, and FFR are all statistically significant. While the second model had Valueonexchanges and ofBTCs as statistically significant, it proved that supply and demand had been the most prominent influences on bitcoins price since its inception.

The primary limitation of these models is the existence of omitted variables. The pictures below show the results of the RAMSEY Reset test, which concludes that this model has omitted variables. Some of these variables could be more specific data relating to the Bitcoin network or other variables related to financial institutions, media, or regulation.

Ramsey Test

Table 3:

```
Ramsey RESET test for omitted variables
Omitted: Powers of independent variables

H0: Model has no omitted variables

F(24, 62) =  15.00
 Prob > F = 0.0000
```

I tested equation 3 for omitted variables using Ramsey Test. The model's functional form is correctly described if there are no omitted variables, which is the null hypothesis.

The alternative theory holds that the model's functional form needs to be correctly described because there are omitted variables.

The test statistic in this instance is an F-statistic with 24 and 62 degrees of freedom, and the likelihood that the null hypothesis would produce an F-statistic that is as extreme as the one observed (or more extreme) is extremely low (0.0000). As a result, we reject the null hypothesis and conclude that the model's functional form is incorrect and that variables have been excluded from the model.

Table 4:

```
. estat ovtest, rhs

Ramsey RESET test for omitted variables
Omitted: Powers of independent variables

H0: Model has no omitted variables

F(6, 1491) = 1804.85
  Prob > F =  0.0000
```

I tested equation 4 for omitted variables using Ramsey Test. The model has no omitted variables, which is the null hypothesis. The test looks for any missing variables that may be added to the model to enhance its fit.

The output displays a p-value of 0.0000 and an F-statistic of 1804.85. We reject the null hypothesis because the p-value is less than the significance level of 0.05, which means that the model may have excluded factors that would have improved its fit. This shows that in order to enhance the model's performance, more model specifications and diagnostics are required.

Breusch-Pagan Test

Below shows the results of the Breusch-Pagan Test, which tests for heteroscedasticity.

Table 5

```
. estat hettest

Breusch–Pagan/Cook–Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of MrktPrice

H0: Constant variance

    chi2(1) =   0.19
Prob > chi2 = 0.6667
```

The alternative hypothesis is heteroskedasticity (variance that is not constant), while the null hypothesis is that the error terms have constant variance.

The test was run on the variable MrktPrice, with the underlying premise being that the error terms are normally distributed. The probability of seeing a test statistic as extreme as the p-value represents the one calculated from the data under the null hypothesis. The test statistic is chi-squared with one degree of freedom.

In this instance, the p-value is 0.6667, and the test statistic is 0.19. I fail to reject the null hypothesis because the p-value is higher than the significance level of 0.05. The data do not contain any evidence of heteroskedasticity.

The p-value is 0.6667, and the test statistic (chi-squared) is 0.19 with 1 degree of freedom. We fail to reject the null hypothesis since the p-value is higher than the significance level, typically 0.05 and conclude there is no meaningful evidence of heteroskedasticity. This shows that the error terms' variance is consistent across all observations and that the homoskedasticity presumptions are probably accurate.

Table 6:

```
Breusch–Pagan/Cook–Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of Price

H0: Constant variance

    chi2(1) = 376.61
Prob > chi2 = 0.0000
```

The above test is for the variable "Fitted values of Price" for heteroskedasticity. The alternative hypothesis is heteroskedasticity, while the null hypothesis is that the error terms have constant variance.

The test statistic is 376.61 and is a chi-squared with one degree of freedom (chi2(1)). The p-value is extremely low (Prob > chi2 = 0.0000), indicating that the null hypothesis is strongly refuted. As a result, the error terms show indications of heteroskedasticity. Therefore, this result suggests that there is heteroskedasticity in the regression model.

Additional weaknesses of this study include not combining the two data sets, as I need help finding an overlapping variable.

Further, there are certain variables that I still need to include. For example, market news plays a significant role in terms of price discovery. However, I needed help to develop an effective metric to measure the market. While market news is roughly included in this equation by including Google trend analysis and tracking trading volume, which is highly correlated with market news, it is not directly included.

Regulatory changes also are less than I would like. I did not devise an effective way to measure any regulatory change or announcement that may affect Bitcoin. I included interest-rate tracking, a powerful tool the United States federal government uses that affect Bitcoin's price. However, any governmental change or hostility towards Bitcoin would not be included. An example would be the Chinese government attempting to ban all bitcoin inside their country, which caused a significant price decrease.

I also could not develop a good descriptor for the estimated time to halving. The Bitcoin halving occurs every 210,000 blocks, massively shifting the price as miners' reward decreases by half. Due to the unpredictable nature of when that halving occurs, I could not come up with an effective means to measure it.

This model should be used to understand further what factors affect the price of Bitcoin. It has clickable use in terms of investment decisions. It can also help people better understand the risk of acquiring or holding Bitcoin. Additionally, this model can aid in further research and development of better models that predict the prices of commodities, specifically Bitcoin.

Variables Appendix

Data Set 1

| **Variable Name** | **Variable definition** | **Source** |
|---|---|---|
| AvgBlockSize | The average block size per day in megabytes | https://www.blockchain.com/explorer/charts/avg-block-size |
| Difficulty | A relative measure of how difficult it is to mine a new block for the blockchain per day | https://www.blockchain.com/explorer/charts/difficulty |
| FFR | The interest rate at which depository institutions trade federal funds (balances held at Federal Reserve Banks) with each other per day | https://fred.stlouisfed.org/series/FEDFUNDS |
| FeePerTransaction | The daily average transaction fees in USD per transaction | https://www.blockchain.com/explorer/charts/fees-usd-per-transaction |
| HashRate | The estimated number of terahashes per second the bitcoin network is performing per day | https://www.blockchain.com/explorer/charts/hash-rate |

| | | |
|---|---|---|
| MarketCap | The measure the size and value of a cryptocurrency equal to price times volume | https://data.nasdaq.com/data/BCHAIN/MKTCP-bitcoin-market-capitalization |
| MarktPrice | The average USD market price across major bitcoin exchanges | https://www.blockchain.com/explorer/charts/market-price |
| NTransactionsperblock | The total number of confirmed transactions per day | https://www.blockchain.com/explorer/charts/n-transactions |
| NVT | Network Value to Transactions. A measure of network activity determined by dividing the Network Value (= Market Value) by the total transactions volume in USD per day | https://www.blockchain.com/explorer/charts/nvt |
| NWalletperd | The total number of unique addresses used on the blockchain | https://www.blockchain.com/explorer/charts/n-unique-addresses |

| | | |
|---|---|---|
| NetworkControlScore | The average percentage of hash power provided to the network by each of the 5 top 5 mining pools | https://www.blockchain.com/explorer/charts/n-unique-addresses |
| TradeVol | The total number of transactions on the blockchain | https://www.blockchain.com/explorer/charts/n-transactions-total |
| VIX | VIX measures market expectation of near-term volatility conveyed by stock index option prices per day | https://fred.stlouisfed.org/series/VIXCLS |
| VolumeBTC | The total number of bitcoins on the Gemini Exchange at the end of a day | https://www.cryptodatadownload.com/data/gemini/ |
| VolumeUSD | The total dollar amount on the Gemini Exchange at the end of a day | https://www.cryptodatadownload.com/data/gemini/ |
| WebSearchesPrecentile | This is a adjusted metric based off of google trends data regarding bitcoin. The data is adjusted from monthly to be a daily | https://trends.google.com/trends/explore?date=all&geo=US&q=bitcoin&hl=en-US |

| | | |
|---|---|---|
| | average from that month. It is a percentile. | |
| Close | This is the daily price that bitcoin was at on the Gemini Exchange at 6pm EST | https://www.cryptodatadownload.com/data/gemini/ |
| Date | This was the day each piece of date was from. This enables the combination of multiple sources of data. | |
| High | This is the highest price that bitcoin was priced at on the Gemini Exchange during each day | https://www.cryptodatadownload.com/data/gemini/ |
| Low | This is the lowest price that bitcoin was priced at on the Gemini Exchange during each day | https://www.cryptodatadownload.com/data/gemini/ |
| ntransactionsexcludingpopular | The total number of transactions excluding those involving the | https://www.blockchain.com/explorer/charts/n-transactions-excluding-popular |

| | | |
|---|---|---|
| | network's 100 most popular addresses | |
| Open | This is the daily price that bitcoin was at on the Gemini Exchange at 4am EST | https://www.cryptodatadownload.com/data/gemini/ |
| Symbol | This is the link | https://www.cryptodatadownload.com/data/gemini/ |

Data Set 2

| **Variable Name** | **Variable definition** | **Source** |
|---|---|---|
| MarktCap | The total USD value of bitcoin in circulation equal to price times volume | https://www.blockchain.com/explorer/charts/market-cap |
| Price | MrktCap/ofBTCs | |
| Valueonexchanges | The total USD value of trading volume on major bitcoin exchanges | https://www.blockchain.com/explorer/charts/trade-volume |
| ofBTCs | The total number of bitcoins that have been mined at any given point | https://www.blockchain.com/explorer/charts/total-bitcoins |
| x | Unique block identifier | |

Notes:

The first and most apparent variable involving Bitcoin is time. The first way to measure this is by using dates and market opening and closing; however, Bitcoin does not open and close like traditional markets. I.e., bitcoin still trades on the weekends. The other problem with this is due to the way bitcoins price fluctuates. Its price is not necessarily set at any given minute since the blockchain could still process the next block at a specific time.

Additionally, different exchanges price Bitcoin differently; there are slight price fluctuations based on which exchange you are looking at. There is also an interesting argument made in the article “Bitcoin is Time,” which suggests that it may be more prudent to use block height as opposed to a specific time on the clock in this analysis.[5] That being said, I do believe date and time is better for this analysis, as there are certain months out of the year, that bitcoin is known to perform better during; for example, April is always a notably good month for bitcoin, while September has traditionally been a very bad month.

[5] https://dergigi.com/2021/01/14/bitcoin-is-time/

References

Ciaian, Pavel & Rajcaniova, Miroslava & Kancs, d'Artis. (2016). The Economics of BitCoin Price Formation. Applied Economics. 48. 1799-1815. 10.1080/00036846.2015.1109038.

Blockchain.com. (n.d.). Market Cap. Retrieved from https://www.blockchain.com/explorer/charts/market-cap

Blockchain.com. (n.d.). Trade Volume. Retrieved from https://www.blockchain.com/explorer/charts/trade-volume

Blockchain.com. (n.d.). Total Bitcoins. Retrieved from https://www.blockchain.com/explorer/charts/total-bitcoins

Blockchain.com. (n.d.). Average Block Size. Retrieved from https://www.blockchain.com/explorer/charts/avg-block-size

Blockchain.com. (n.d.). Difficulty. Retrieved from https://www.blockchain.com/explorer/charts/difficulty

Federal Reserve Bank of St. Louis. (n.d.). Effective Federal Funds Rate [FEDFUNDS]. Retrieved from https://fred.stlouisfed.org/series/FEDFUNDS

Blockchain.com. (n.d.). Fees USD Per Transaction. Retrieved from https://www.blockchain.com/explorer/charts/fees-usd-per-transaction

Blockchain.com. (n.d.). Hash Rate. Retrieved from https://www.blockchain.com/explorer/charts/hash-rate

CryptoDataDownload. (n.d.). Gemini Bitcoin Trade Data. Retrieved from https://www.cryptodatadownload.com/data/gemini/

Blockchain.com. (n.d.). Market Price. Retrieved from https://www.blockchain.com/explorer/charts/market-price

Blockchain.com. (n.d.). Number of Transactions. Retrieved from https://www.blockchain.com/explorer/charts/n-transactions

Blockchain.com. (n.d.). Network Value to Transactions Ratio (NVT). Retrieved from https://www.blockchain.com/explorer/charts/nvt

Blockchain.com. (n.d.). Number of Unique Addresses. Retrieved from https://www.blockchain.com/explorer/charts/n-unique-addresses

Blockchain.com. (n.d.). Total Number of Transactions. Retrieved from https://www.blockchain.com/explorer/charts/n-transactions-total

Federal Reserve Bank of St. Louis. (n.d.). CBOE Volatility Index: VIX [VIXCLS]. Retrieved from https://fred.stlouisfed.org/series/VIXCLS

Google Trends. (n.d.). Bitcoin. Retrieved from https://trends.google.com/trends/explore?date=all&geo=US&q=bitcoin&hl=en-US

Blockchain.com. (n.d.). Number of Transactions Excluding Popular Addresses. Retrieved from https://www.blockchain.com/explorer/charts/n-transactions-excluding-popular

Nasdaq. (n.d.). Bitcoin Market Capitalization. Retrieved from https://data.nasdaq.com/data/BCHAIN/MKTCP-bitcoin-market-capitalization

Coinbase. (n.d.). Coinbase Data. Retrieved from https://data.coinbase.com/