CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Exploring Bitcoin Volatility

Advanced Econometrics Project

Authors: David Černý, Michael Tor
Study program: Economics and Finance

Year: 2024

Abstract

This study investigates Bitcoin volatility and its drivers between the years 2020 and 2023. We developed a simple regression model of the volatility (estimated by Garman-Klass estimator), which was further estimated using an IV (Instrumental Variable) approach, due to the assumptions of endogeneity of price. We discover a positive link between Bitcoin volatility and traded volume in $USD and another robust association is identified with Google searches for the keyword "bitcoin". Our findings emphasize the significance of public interest and its proxies as primary factors driving Bitcoin volatility. This research contributes to the literature on the complex dynamics shaping Bitcoin's volatility while using the most up-to-date data.

Keywords Bitcoin, Volatility, Google Searches, Endogeneity

Title Exploring Bitcoin Volatility

Contents

Chapter 1

Introduction and Literature review

1.1 Brief Introduction

A popular and dynamic financial asset in recent years, Bitcoin, has drawn interest from investors, economists, and the general public. Bitcoin's decentralized characteristics and its potential to transform established banking systems caught the eye of almost everyone. Investment in cryptoassets is distinct because of its tremendous volatility and sharp swings, which are offset by the possibility of significant price declines. The Bitcoin is most successful example. Users can conduct digital transactions globally through the peer-to-peer decentralized network formed by the open-source Bitcoin protocol (Nakamoto 2008).

Understanding the variables behind Bitcoin price volatility is becoming increasingly important for both market participants and policymakers as the cryptocurrency market continues to develop ever so quickly.

This analysis aims to explore the factors affecting volatility in Bitcoin prices and offer a viewpoint on the subject by using up-to-date data.

This project seeks to contribute to the growing literature by empirically investigating the newest data possible in the 2020 to 2023 time frame.

Our motivation for this project is the increasing significance of Bitcoin in the global financial market and the need to understand the functionality underlying its volatility. Ultimately, to make educated decisions and manage risk in this situation, it is essential to comprehend the causes of the volatility in Bitcoin prices and what affects them.

Interest rates and macroeconomic indicators all have an impact on traditional financial assets, such as stocks. However, the bitcoin market behaves to some extent differently, being more influenced by sentiment, perception, and information flow in addition to basic economic principles.

Furthermore, the Bitcoin market is an interesting subject for empirical research because of its quick evolution, popularity, and data availability. We therefore intend to advance knowledge of the complex processes at work in digital asset markets.

1.2 Literature review

The first who examined the price dynamics of bitcoin was Kristoufek (2013), who found the significant impact of the amount of Google searches and number of Wikipedia queries as explanatory variables, variables which served as a proxy for public interest.

Studying the literature about the volatility of Bitcoin, Bystrom & Krygier (2018) made use of monthly, weekly, and daily data from 2011 to 2017. And found a small but positive contemporaneous relationship between changes in the volatility of Bitcoin and changes in the volatility of the trade-weighted USD currency index, shown by correlations and regressions. They also found, that there is a more substantial positive correlation between the volatility of Bitcoin and the volume of searches on Google for terms linked to Bitcoin, especially the word "bitcoin." The only factors that predict future Bitcoin volatility are Google searches for the phrase "bitcoin" and, to a lesser extent, the volatility of the USD currency index.

Estrada (2017) in his analysis of volatility, found no statistically significant Granger-causality when analyzing the time series of BTC and the S&P 500 and BTC and the VIX. These results suggest Bitcoin price has no causal relationship to financial instruments such as S&P 500. He concludes, that it cannot be ruled out that the Bitcoin weekly price does not correlate with Blockchain trends or that Bitcoin realized volatility does not correlate with the S&P 500.

Both Kubal & Kristoufek (2022); Černý (2023) investigated the price dynamics of bitcoin using a system of simultaneous equations. The results indicate, that the endogeneity is omnipresent in the system and thus must be tackled, which is performed using the IV approach. Price among other things was found to be most likely endogeneous.

Chapter 2

Data and Methodology

This Chapter begins with a description of the data used and the intuition behind the selection of the variables we chose. Then at the end of the chapter we describe the statistical methods used for our analysis.

2.1 Data Description

As was mentioned in Chapter 1, the goal of this paper is to explore factors, which drive the price volatility of Bitcoin. Since Bitcoin could be considered an emerging environment and the relationships between variables in such a system might rapidly change, we decided to aim at the more recent time frame. Thus, we use daily data from January 1, 2020, to December 31, 2023, i.e. the period that may be defined first by a quick market downturn due to COVID-19, and then by an ongoing bull market, which escalated to bitcoin All Time High (ATH) price in 2021, which was followed by a gradual drop in price from which bitcoin did not recover sufficiently even in 2023.

The data collection process could be narrowed to two sources: Data obtained from **Yahoo Finance**[1] and data from **Google Trends**[2].

Yahoo Finance

Most of the data was imported from Yahoo Finance, using a pipeline written in R. These data are:

Daily **OHLC**[3] **Prices** of bitcoin, which were subsequently used to estimate

[1] `https://finance.yahoo.com/`
[2] `https://trends.google.com/trends/`
[3] Opening, highest, lowest, and closing prices during given timeframe

Figure 2.1: Bitcoin Close Price

volatility using Garman & Klass (1980) estimator, using the formula:

$$\sigma_t = \sqrt{\frac{1}{2} * \left[\log\left(\frac{h_t}{l_t}\right)\right]^2 - [2\log 2 - 1] * \left[\log\left(\frac{c_t}{o_t}\right)\right]^2} \tag{2.1}$$

which serves as a dependent variable in our model.

Traded volume in $USD, which sums the size of all trades occurring in a given day.

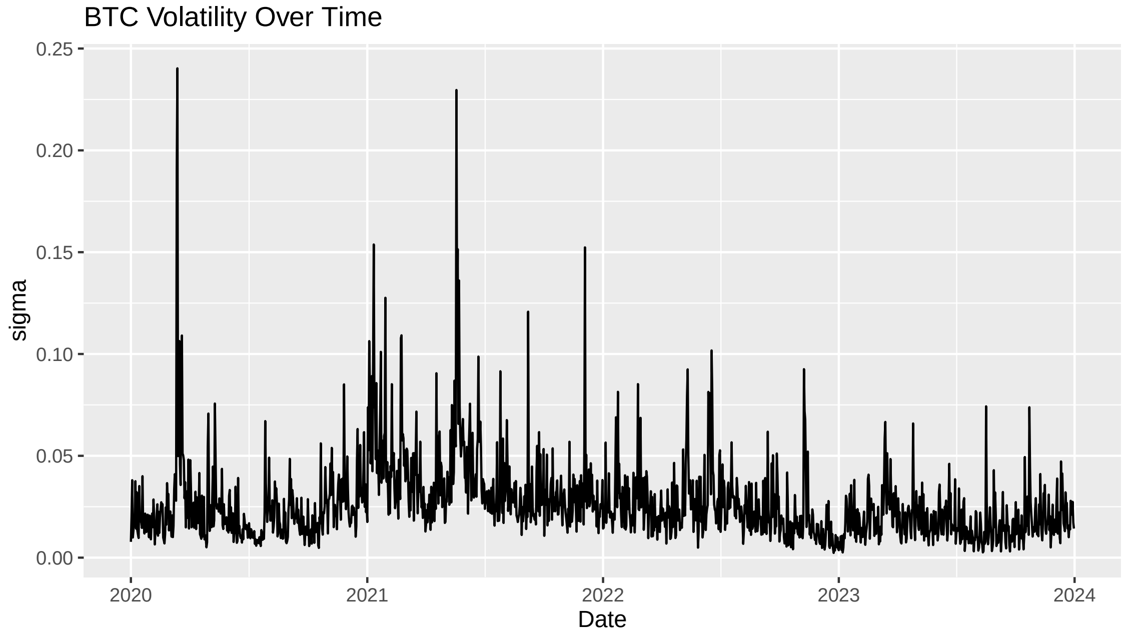
S&P 500: Daily close prices of the Standard and Poor's 500 (S&P 500) index, which serves as a proxy for global economic development.

VIX: Daily close values of the Chicago Board Options Exchange's CBOE Volatility Index (VIX), which serves as a proxy of volatility in standard financial markets.

Google Searches

The last explanatory variable we use is the frequency of Google searches for the keyword *bitcoin*, which was obtained using a python script connecting Google Trends API. Google searches are popular explanatory variable for bitcoin price since Kristoufek (2013). We believe that it might play a vital role in the determination of bitcoin volatility since it serves as a proxy for public interest and the increase in public interest might increase speculation on the price and thus also volatility.

Figure 2.2: Bitcoin Price Volatility



2.2 Methodology

We propose the simple linear regression model:

$$\begin{aligned}\log(sigma_t) = \alpha_0 + \alpha_1 \log(close_t) + \alpha_2 \log(vix_t) + \alpha_3 \log(volume_t) \\ + \alpha_4 \log(google\_searches_t) + \epsilon_t\end{aligned} \tag{2.2}$$

which, assuming all variables are exogenous, can be estimated using the Ordinary Least Squares (OLS) approach. However, we can assume that the price is not exogenous and is determined by other variables, which is indicated by Kubal & Kristoufek (2022) or Černý (2023).

Exogeneity is a property, which means that the explanatory variable is not correlated with residuals and is a necessary condition for the OLS estimator to be unbiased. If some of our variables are exogenous, we can deal with it using several approaches. One such approach is **Instrumental Variable** (IV) regression, where we find variable z, which is **relevant** in the system of endogenous variable x ($Cov(z, x) \neq 0$) and is also **exogeneous** in the initial system ($Cov(z, \epsilon) = 0$). In such case, we can estimate the associated estimator of β using

$$\hat{\beta} = \frac{Cov(z, y)}{Cov(z, x)}$$

where y is a dependant variable (Wooldridge 2015).

To decide whether to use OLS or IV approach we use Hausman (1978) test with H_0 : Both approaches are consistent but OLS is more efficient and H_A : Only IV is consistent.

Chapter 3

Results and Discussion

This chapter provides an in-depth exploration of the entire modeling process, including the selection of suitable econometric methods, statistical testing, results presentation, and the implications that follow.

Model Selection

The first obstacle, as outlined in chapter 2, we encountered was the decision, of whether to use standard OLS or endogeneity robust approach. We firstly estimated the model using both OLS and IV approaches, using S&P500 as an instrument and tested their consistency using Hausman (1978) test. We obtain the Hausman test statistics of 25.61, which implies the p-value < 0.001, thus we can reject the H_0 : OLS is consistent, and proceed with IV model.

Robustness testing

After the model estimation, we tested the robustness of our estimator aiming to check for possible violations of model assumptions. We tested the following violations:

1. We started with the testing of the unit root of residuals. We performed the Augmented Dickey-Fuller (2009) (ADF) test, with H_0 : residuals containing unit root, which is for our purposes crucial to reject. In the testing, we rejected the ADF test at any reasonable significance level.

2. Furthermore we tested the possible heteroscedasticity and autocorrelation of residuals. Using Breusch & Pagan (1979) test for heteroscedasticity and Durbin & Watson (1950) test for autocorrelation we were not able

to reject either heteroscedasticity or autocorrelation of residuals, which indicates that our estimator is not the most efficient.

Therefore we opted to use heteroscedasticity and autocorrelation consistent standard errors (HAC), to obtain valid standard errors (MacKinnon & White 1985).

Results

The relationship between variables of interest is presented in the following table:

Table 3.1: IV estimator.

	Estimate	Std. Error	t-statistic	Pr(>|t|)
Drivers of Bitcoin volatility:				
(Intercept)	-19.724	1.720	-11.465	<0.001 ***
log(close)	-0.014	0.067	-0.211	0.833
log(vix)	0.039	0.085	0.457	0.648
log(google_searches)	0.472	0.095	4.950	<0.001 ***
log(volume)	0.619	0.061	10.149	<0.001 ***
		R^2	Adj. R^2	
		0.593	0.592	
Number of observation				1461
Note:	*p<0.1; **p<0.05; ***p<0.01			

Our model seems to explain almost 60% of the variance in the volatility, with Adjusted R^2 being equal to 0.592.

It can be seen that only 2 variables were discovered to be statistically significant. There was not found any significant effect of the bitcoin price on the volatility. In the case of VIX, it is indicated that the bitcoin volatility is not driven by the volatility in standard financial markets as might be expected.

Google searches turned out to have a significant positive effect on the volatility, which is in accordance with the results from Bystrom & Krygier (2018). This indicates that the higher the public interest is the greater the volatility will be, which is in line with our expectations.

Furthermore, the traded volume of $USD is likely to have a significantly positive effect on the price volatility of Bitcoin, with an elasticity of approximately 62%. To our best knowledge, the volume of $USD is not usually used as an explanatory variable for volatility modeling, therefore these results might be surprising. We interpret these results that the greater traded volume indicates more activity in the network, which could bring more speculation and similarly to the effect of the Google searches could increase volatility.

Chapter 4

Conclusion

In conclusion, due to its rapid evolution, the bitcoin market is an intriguing topic for empirical analysis. This econometrics project looks into the connection between public interest activity and Bitcoin volatility in an effort to better understand the complex processes at play in digital financial asset markets.

We aimed to explore the dynamics of bitcoin price volatility. We examined daily data between the years 2020 and 2023 and since we assumed endogeneity of closing price we employed the Instrumental Variable (IV) regression model.

There are several interesting findings. The results indicate that the bitcoin price volatility is not affected by volatility in standard financial markets. Furthermore, we find out the positive effect of Google searches and $USD traded volume, variables which might be considered as proxies for public interest.

Bibliography

Breusch, T. S. & A. R. Pagan (1979): "A simple test for heteroscedasticity and random coefficient variation." *Econometrica: Journal of the econometric society* pp. 1287–1294.

Bystrom, H. & D. Krygier (2018): "What drives bitcoin volatility?" *Available at SSRN 3223368* .

Černý, D. (2023): "Impact of total transaction fees on the price of bitcoin and ethereum." .

Durbin, J. & G. S. Watson (1950): "Testing for serial correlation in least squares regression: I." *Biometrika* **37(3/4)**: pp. 409–428.

Estrada, J. C. S. (2017): "Analyzing bitcoin price volatility." *University of California, Berkeley* .

Fuller, W. A. (2009): *Introduction to statistical time series.* John Wiley & Sons.

Garman, M. B. & M. J. Klass (1980): "On the estimation of security price volatilities from historical data." *Journal of business* pp. 67–78.

Hausman, J. A. (1978): "Specification tests in econometrics." *Econometrica: Journal of the econometric society* pp. 1251–1271.

Kristoufek, L. (2013): "Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era." *Scientific reports* **3(1)**: p. 3415.

Kubal, J. & L. Kristoufek (2022): "Exploring the relationship between bitcoin price and network's hashrate within endogenous system." *International Review of Financial Analysis* **84**: p. 102375.

MacKinnon, J. G. & H. White (1985): "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties." *Journal of econometrics* **29(3)**: pp. 305–325.

Nakamoto, S. (2008): "Bitcoin: A peer-to-peer electronic cash system." *Decentralized business review* p. 21260.

Wooldridge, J. M. (2015): *Introductory econometrics: A modern approach.* Cengage learning.

Appendix A

Additional materials

All the necessary material to replicate the work, together with an python script to retrieve data from google trends, data and main R script is publicly available and easily accessible at `https://github.com/cedav12/Exploring_Bitcoin_Volatility/tree/main`