Data Collection and Preprocessing :

Data Overview :

For measuring investor sentiment in bitcoin in this study three proxies are used .

1. Wikipedia searches about bitcoin:
2. Google trends searches about bitcoin :
3. Fear and Greed index :
4. Bitcoin Trading Volatility Index

Data Collection :

Investor attention proxies: for modelling investor sentiment for bitcoin wikipedia searches , google trends searches about bitcoin and fear and greed index was used .

Wikipedia Proxy : Utilize Wikipedia search trends as a proxy to analyze the public interest and awareness of Bitcoin, as these could give insights to the temporal variation of public curiosity - especially in reactions to major events in price movements in the history of Bitcoin. The Wikipedia API was then used to gather data programmatically from the metrics on the site.

The Wikipedia API (`https://wikimedia.org/api/rest_v1/`) was used to extract search volume data for the term "Bitcoin" across all languages. This ensures a global perspective on public interest.
The time period for which data was collected was from 2015 to the year 2024.

Google trends proxy : Google Trends data was used to estimate public interest in Bitcoin over time. Google Trends is an extremely useful measure of search behavior; it measures how often one term is queried on Google relative to the total search volume for a given period of time.
One thing to keep in mind is that initially data of trends collected was relative to the time period . Therefore in the later part of the study it would be shown how it was converted on an absolute scale over the time period from 2015 to year 2024 .

Fear and Greed index : To quantify the psychological forces driving Bitcoin's market, the Fear and Greed Index was incorporated into this research. An index commonly utilized in financial markets, the Fear and Greed Index measures investors' emotions of extreme fear and extreme greed with respect to numerous quantitative and qualitative metrics. The Fear and Greed Index provides Bitcoin with a sentiment proxy in its cryptocurrency market. It offers information on the manner in which these psychological factors correlate with price trends and public interest in the market.

The concerned data was collected from alternative.me.

Financial Variables related to Bitcoin : In this study the financial variables used were Volume ,Opening Price , Closing Price .
Volume : Volume tells the number of stock of a company traded during a day .
All these variables were extracted from Yahoo finance .

Feature engineering :

"In this section, a new feature, the Bitcoin Trading Volatility Index, was derived from the existing feature, *Volume*. The Bitcoin Trading Volatility Index is calculated as the logarithmic ratio of the current day's volume to the previous day's volume, offering a measure of day-to-day fluctuations in Bitcoin trading activity.

Wikipedia search index was created from existing feature wiki searches . The Wikipedia search index is calculated as the logarithmic ratio of the current day's wiki searches of bitcoin to the previous day's wiki search , offering a measure of day-to-day fluctuations in wikipedia searches.

Google trends Data Normalisation : Google Trends data is scaled relative to the highest search volume within a given time frame, so scales across different intervals (0–100) are independent and not comparable. Normalizing one dataset to the other's scale ensures consistency and allows for meaningful comparisons.
Steps to Normalize Using an Overlapping Period

1. Identify the Overlapping Period:
   Determine the common time frame where both datasets overlap.
- Example:
    - Dataset 1: January 2015–December 2020
    - Dataset 2: January 2020–December 2025
    - Overlapping period: January 2020–December 2020
2. Calculate the Scaling Factor:

   - Extract overlapping period data from both datasets.
- Calculate the scaling factor:

Scaling Factor = Average Popularity (Dataset 1)/Average Popularity (Dataset 2)

3. Use the Scaling Factor:
   Adjust the values in Dataset 2 according to Dataset 1 scale:
Normalized Popularity (Dataset 2) = Popularity (Dataset 2) *{Scaling Factor}

For this study data was collected in chunks keeping in mind above constraints so that in those consecutive chunks an overlap could be found so that data normalization could be done .

Google Search index : Google Search index was created from existing feature normalised trends searches . The Google Search index is calculated as the logarithmic ratio of the current day's trends searches to the previous day's trends search , offering a measure of day-to-day fluctuations in trends searches about bitcoin.

Bitcoin Returns : Bitcoin Returns was created from the existing feature Bitcoin Closing Price. The Bitcoin Returns is calculated as the Percent change in bitcoin closing price .

Methodology : For measuring volatility for each proxy we have employed several Garch models . These include The GARCH (1) model , The EGARCH (1) model, The GJR-GARCH (1) model, The AP-ARCH (1) model .

1. The GARCH (1) model : (Generalized Autoregressive Conditional Heteroskedasticity): The GARCH(1,1) model is a standard volatility model that estimates current variance as a function of past variance (autoregressive term) and past squared returns (moving average term).
   It is extensively used in the capture of persistent volatility and clustering effects of financial time series where periods of high volatility are usually followed by periods of high volatility and vice versa.

   $$r_t = \mu + \Lambda_1 S_{t-1} + \varepsilon_t$$

   $$\sigma_t^2 = \omega + \alpha\left(\varepsilon_{t-1}^2\right) + \beta\left(\sigma_{t-1}^2\right) + \Lambda_2 S_{t-1}$$

2. The Egarch(1) model : (Exponential GARCH):
   The EGARCH model is a generalization of GARCH to allow for asymmetric effects of shocks.Unlike GARCH, it models the logarithm of variance, which means that variance is always positive without requiring any constraints on parameters.It captures the leverage effect where negative and positive shocks produce volatility of different magnitude .

   $$ln\left(\sigma_t^2\right) = \omega + \alpha\left(|z_{t-1}| - E|z|\right) + \beta\left(\sigma_{t-1}^2\right) + \gamma(z_{t-1}) + \Lambda_2 S_{t-1}$$

   Gamma represents asymmetry here meaning leverage effect . It signifies that the market responds differently to positive and negative shocks .If gamma =0 it represents symmetry. If gamma < 0 negative shocks will increase the volatility more than positive shocks. If gamma > 0 positive shocks increase the volatility more than negative shocks .

3. GJR-GARCH(1,1) (Glosten-Jagannathan-Runkle GARCH):
   The GJR-GARCH model builds on the GARCH framework by incorporating asymmetry through a dummy variable.It distinguishes between positive and negative shocks, allowing for an explicit modeling of the leverage effect.

   $$\sigma_t^2 = \omega + (\alpha + \gamma I_{t-1})\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \Lambda_2 S_{t-1}$$

If gamma has no asymmetric volatility. If gamma > 0 negative shocks will increase the volatility more than positive shocks. If gamma < 0 positive shocks increase the volatility more than negative shocks.

Results :

For the purpose of modeling volatility, the GARCH model is excluded from the analysis, as the primary focus is on capturing asymmetries in the data.

1. Wiki-Search-Index :

Model	AIC	BIC	Gamma	Gamma-P-Value
Egarch	18212.3	18255.3	-0.0407	0.200
GJR-GARCH	18234.3	18277.4	5.583e-02	0.316
Garch	18235.4	18272.3	Perfectly symmetric	

From the above table it is concluded that the E-GARCH model is best and gamma representing asymmetry is insignificant .

2. Google-Search-index:

Model	AIC	BIC	Gamma	Gamma-P-Value
Garch	19008.8	19046.0	Perfectly symmetric	
E-Garch	18986.2	19029.6	-0.0420	0.153

GJR-GARCH	19008.1	19051.5	0.0572	0.298

From the above table it is concluded that the E-GARCH model is best and gamma representing asymmetry is insignificant .

3. Bitcoin Trading Volatility Index :

Model	AIC	BIC	Gamma	Gamma-P-Value
Garch	19781.2	19818.6	Perfectly symmetric	
E-Garch	19757.6	19801.3	-0.0395	0.144
GJR-GARCH	19781.5	19825.1	0.0506	0.310

From the above table it is concluded that the E-GARCH model is best and gamma representing asymmetry is insignificant .

4. Fear and Greed proxy :

Model	AIC	BIC	Gamma	Gamma-P-Value
Garch	-9837.94	-9803.02	Perfect symmetry	
E-Garch	-9870.7	-9840.13	-0.0622	0.110
GJR-GARCH	-9876.39	-9835.65	0.0998	0.177

Here it is concluded that E-garch model is best depending upon AIC,BIC also it is concluded that fear and greed proxy is best as due to its low aic score as compared to other proxies as well as due to a significant low p-value of gamma representing asymmetry which is the core focus of this study .

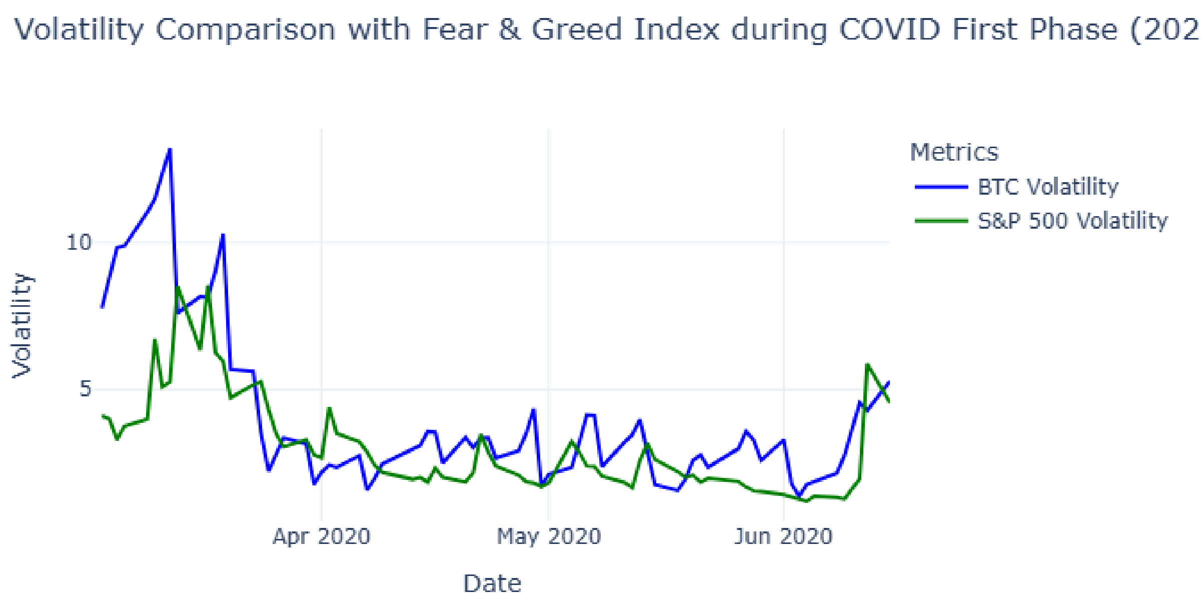
Covid Period Analysis :

In this section of the research, a comparison of volatility analysis between Bitcoin and the S&P 500 index across different stages of the COVID-19 crisis will be undertaken by taking Fear and Greed index as the investor sentiment proxy for both the indices . This study seeks to examine differences in volatility dynamics between these two assets in various stages of the crisis. The crisis will be divided into 5 distinct phases: The First Pandemic Stage,Recovery Phase Post-COVID-19 (2020),Vaccination Phase,Delta Variant Surge,Recovery Period After the Delta Variant.The research will measure volatility as well as asymmetry in both these indices during the above mentioned covid periods using E-GARCH model which has been proven best based on the above work on different proxies . Also

1. The First Pandemic Stage (1 march to 15 june) :

Index	alpha	alpha-P value	beta	Beta p-value	gamma	Gamma p-value
Bitcoin	-0.6611	0.000000e +00	0.8682	0.000e+00	-0.2230	6.905000e -280

S&P 500	0.4053	5.100000e-01	0.8662	4.911000e-08	-0.3257	4.500000e-01



For bitcoin asymmetry is statistically significant whereas for s&p 500 asymmetry is statistically not significant .In the first pandemic stage gamma value for bitcoin is negative meaning the market of bitcoin was more volatile for negative shocks as compared to positive shocks . For bitcoin alpha is statistically significant meaning past shocks have a meaningful impact on current volatility which is absent in s&p 500 . Both bitcoin and s&p 500 have significant beta implying volatility is highly persistent over time for them .

2. Recovery Phase Post-COVID-19 (2020) : July to October 2020

Index	alpha	alpha-P value	beta	Beta p-value	Gamma	Gamma p-value
Bitcoin	0.001317	0.99700	0.000000	1	-0.475	0.071
S&P 500	0.676300	5.704000e-02	0.710600	2.88e-10	-0.059800	6.77e-01

Bitcoin :

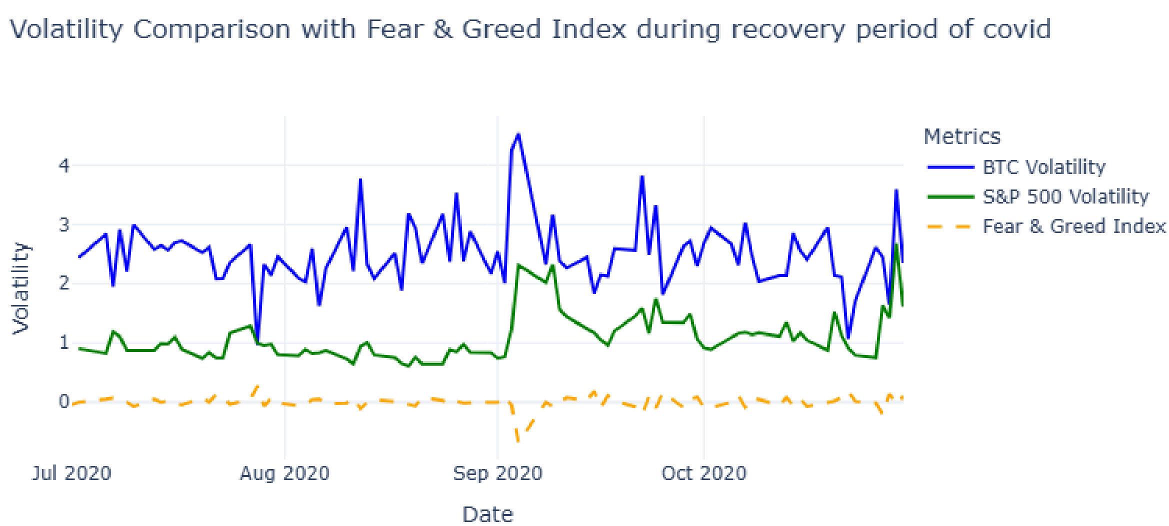
The lack of statistical significance in the **alpha** and **beta** coefficients suggests that past shocks and past volatility have little to no impact on Bitcoin's current volatility. The **gamma** coefficient, though negative, is marginally significant, indicating a possible but weak leverage effect.

These results imply that Bitcoin may not exhibit strong volatility clustering or asymmetric responses to shocks, and its volatility may not be strongly dependent on past market behavior.

S&P 500:

The **alpha** and **beta** coefficients for the S&P 500 are statistically significant, with **beta** showing high persistence of volatility. This indicates that past volatility plays a significant role in determining future volatility, as is typical in traditional markets.

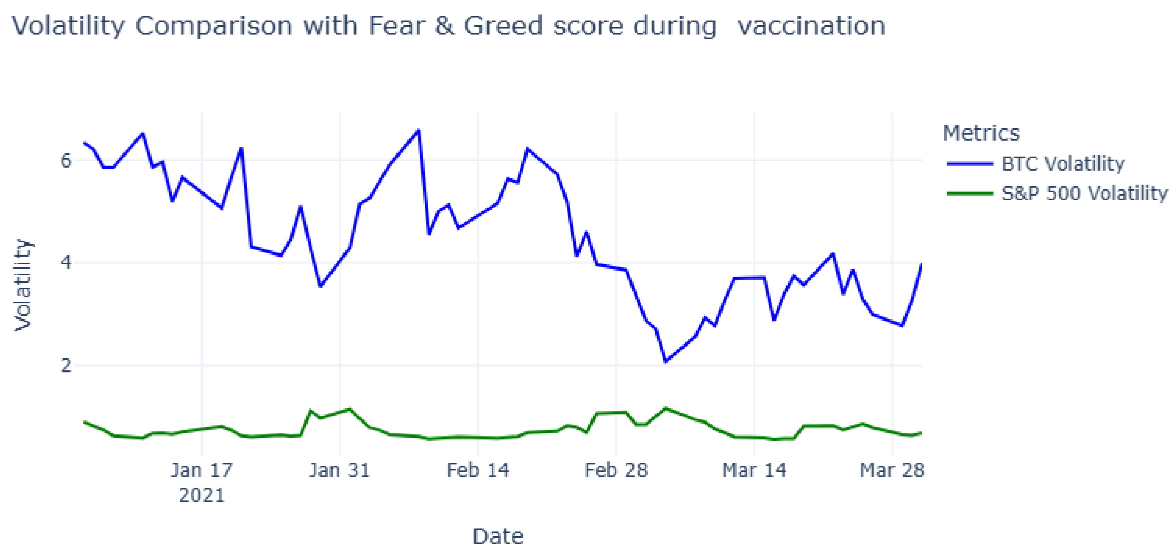
The **gamma** coefficient is insignificant, suggesting that the S&P 500 does not show a significant asymmetric response to positive vs. negative shocks.



3. January to March 2021: Optimism Amid Vaccination Rollouts

Index	alpha	alpha-P value	beta	Beta p-value	gamma	Gamma p-value
Bitcoin	-0.4583	0.0	0.9870	0.0	0.0282	0.0

S&P 500	0.072800	0.974	0.802700	0.451	-0.2036	0.948



Bitcoin :

The **alpha** coefficient is negative and statistically significant (p-value = 0.0). This suggests that past shocks have a significant **inverse relationship** with current volatility.

The **beta** coefficient is very close to 1 and highly significant (p-value = 0.0), indicating a **high level of volatility persistence**

The gamma coefficient is positive and statistically significant implying positive shocks cause more volatility than negative shocks in this period .

S&P 500

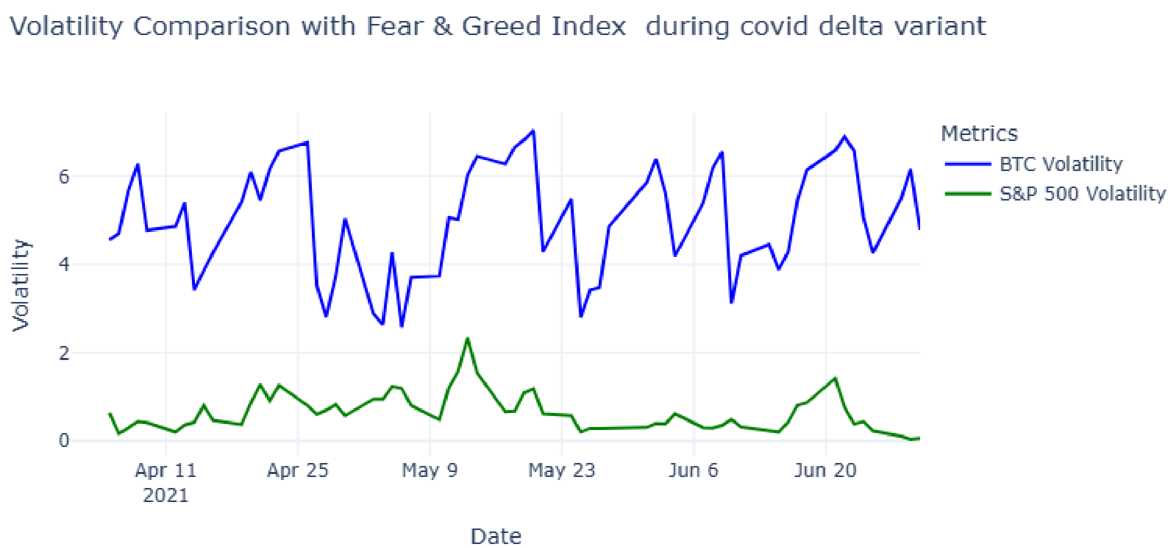
The **alpha** coefficient is positive but **not statistically significant** (p-value = 0.974). This suggests that past shocks do not have a significant impact on the current volatility of the S&P 500.

The **beta** coefficient is positive but **not statistically significant** (p-value = 0.451), indicating that there is no strong persistence in volatility in the S&P 500.

The **gamma** coefficient is negative and **not statistically significant** (p-value = 0.948), indicating that there is no significant leverage effect for the S&P 500.

4. Delta Variant Surge

Index	alpha	alpha-P value	beta	Beta p-value	gamma	Gamma p-value
Bitcoin	-0.4196	0.159000	0.5247	0.006892	-0.4418	0.013940
S&P 500	-0.308800	1.360000e-01	0.800500	1.000000e-89	-0.9683	5.491000e-04



Bitcoin :

The **alpha** coefficient is negative but **not statistically significant** (p-value = 0.159), suggesting that past shocks do not have a significant impact on future volatility for Bitcoin.

The **beta** coefficient is positive and statistically significant (p-value = 0.006892), indicating **moderate volatility persistence**

The **gamma** coefficient is negative and statistically significant (p-value = 0.013940), suggesting the presence of a **leverage effect**. This indicates that negative shocks increase future volatility more than positive shocks.

s&P500 :

The **alpha** is not significant, indicating that past returns do not significantly affect future volatility.

The **beta** is highly significant, suggesting strong **volatility persistence** and indicating that past volatility has a significant impact on future volatility in the S&P 500.

The **gamma** is negative and significant, suggesting the presence of a **leverage effect** in the S&P 500, where negative shocks increase future volatility more than positive shocks.

5. Recovery and regional divergence (july to september 2021):

Index	alpha	alpha-P value	beta	Beta p-value	gamma	Gamma p-value
Bitcoin	-0.6826	2.241000e-02	0.5489	1.105000e-03	-0.1300	2.510000e-01
S&P 500	-0.799	2.970000e-44	0.7925	0.000000e+0	-0.771	1.394000e-35

Bitcoin :

The **alpha** is statistically significant and negative, suggesting that negative past shocks reduce future volatility in Bitcoin.

The **beta** is significant and positive, indicating moderate volatility persistence, meaning that Bitcoin experiences some clustering in volatility, though to a lesser extent than traditional financial markets.

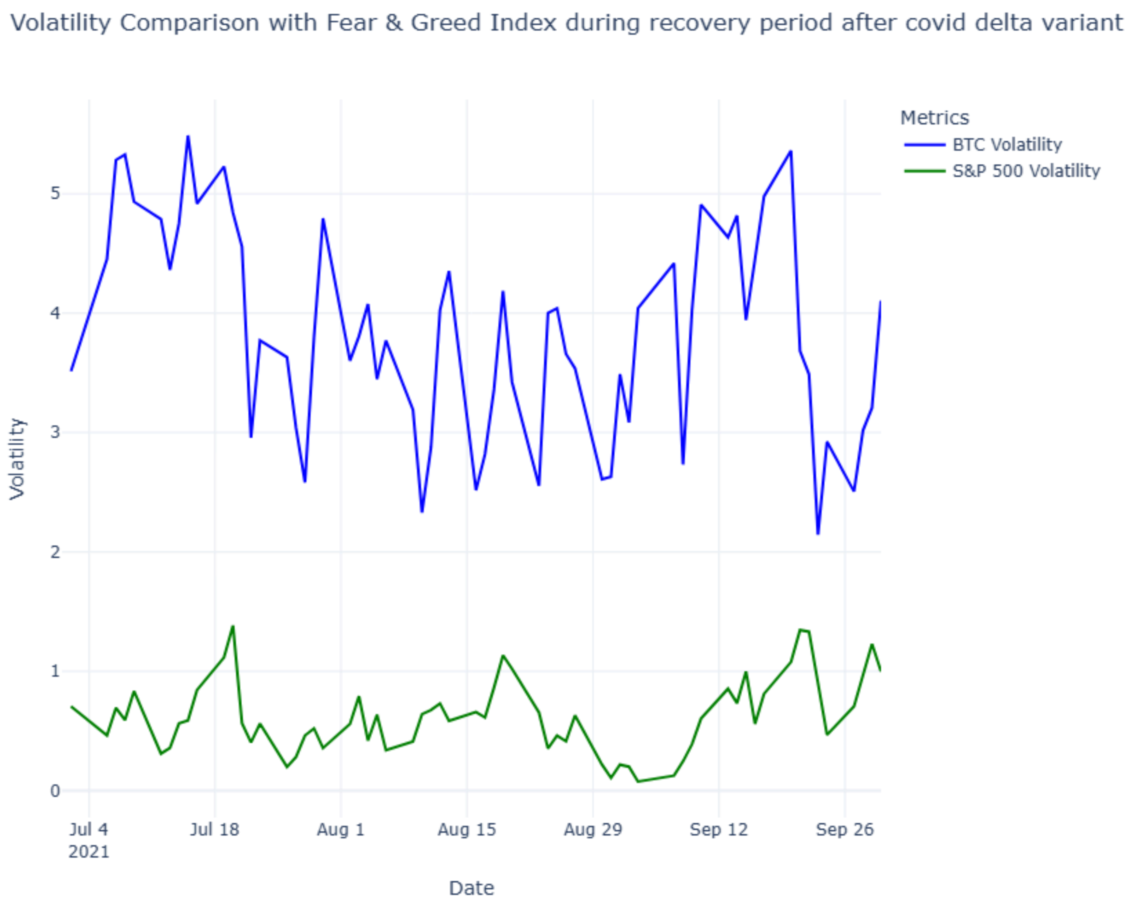
The **gamma** is not significant, suggesting that Bitcoin does not exhibit a significant leverage effect in this model.

s&P 500 :

The **alpha** is highly significant and negative, suggesting that negative shocks reduce future volatility in the S&P 500.

The **beta** is highly significant and positive, indicating strong volatility persistence. This suggests that past volatility in the S&P 500 plays a significant role in determining future volatility.

The **gamma** is highly significant and negative, indicating a strong leverage effect in the S&P 500. Negative shocks lead to higher future volatility than positive shocks of equal magnitude.


Volatility Comparison with Fear & Greed Index during recovery period after covid delta variant
Metrics
BTC Volatility
S&P 500 Volatility
Volatility
0
1
2
3
4
5
Jul 4
2021
Jul 18
Aug 1
Aug 15
Aug 29
Sep 12
Sep 26
Date