# Exploring the Correlation Between Cryptocurrency Price Fluctuations and GitHub Commits: A Web Data Analysis Approach

Cade Wilson

COLUMN Lab/INFO 5871

University of Colorado - Boulder

Dr. Brian Keegan

15 December 2023

**Abstract**

This study delved into the intricate relationship between the fluctuations in cryptocurrency prices and the corresponding impact on the frequency of code contributions by developers engaged in related blockchain projects. To achieve this, a methodical approach was undertaken to gather and analyze historical pricing data of a diverse array of cryptocurrencies, and this data was meticulously extracted from Yahoo Finance, employing web automation techniques. The focus was primarily on the opening and closing prices of these cryptocurrencies over a selected one-year period, with the objective of scrutinizing the nuances of daily price volatility and trying to detect signals which may impact pricing and development.

In parallel, an examination of the development activity was conducted. This involved leveraging GitHub's API to access and compile a comprehensive dataset of commit history metadata such as commit message and time from the principal code repositories of leading cryptocurrencies. Essential details such as the timestamps of the commits and the contents of the commit messages were meticulously recorded and stored for analysis.

The central research question driving this study inquired into the nature of the relationship between the fluctuations in the prices of cryptocurrency tokens, as meticulously reported by Yahoo Finance, and the level of commit activity by developers on the corresponding GitHub repositories dedicated to each specific cryptocurrency. The question at hand was, what are the relationships between GitHub commit activity and market price movements?

To address this question, sophisticated statistical methods were employed to determine whether there was a significant and meaningful predictive relationship between the instability of cryptocurrency pricing and the level of collaboration and activity within the open-source communities involved in cryptocurrency development. This exploration aimed to unravel

whether movements in market prices serve as a deterrent or a catalyst for developer contributions to the diverse and decentralized networks and ecosystems that constitute the cryptocurrency domain. Furthermore, by quantifying and interpreting these dynamics, the study aimed to shed light on the broader factors that are likely to shape the trajectory and evolution of cryptocurrency development in the foreseeable future. This could potentially offer valuable insights into how market forces and developer activities are intertwined in this rapidly evolving and innovative field.

## Research Design

In the study, a comprehensive analysis was conducted to explore the intricate relationship between cryptocurrency price fluctuations and development activity. This approach involved a quantitative and correlational methodology, focusing on historical pricing data for the year of 2023 and development activity metrics. The objective was to identify any significant correlations between the volatility of cryptocurrency prices and the intensity of developer contributions to their respective blockchain codebases.

The research encompassed a carefully selected basket of prominent cryptocurrency tokens, including Bitcoin, Ethereum, Solana, Cardano, Litecoin, Polygon, XRP, Dogecoin, and Binance. Extensive pricing data was collected from Yahoo Finance, covering a full year from December 16th, 2022, to December 16th, 2023; this period was chosen to capture recent trends and fluctuations in the cryptocurrency market. The data extraction process, automated using Python's Selenium, involved collecting daily data points such as opening and closing prices, highs and lows, adjusted closing values, trading volumes, and timestamps. This method ensured a comprehensive dataset for analysis.

| | Date | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|---|
| **0** | 2023-12-16 | 41937.74 | 42071.65 | 41723.11 | 42071.65 | 42071.65 | 19894603776.00 |
| **1** | 2023-12-14 | 42884.26 | 43390.86 | 41767.09 | 43023.97 | 43023.97 | 25578530178.00 |
| **2** | 2023-12-13 | 41468.46 | 43429.78 | 40676.87 | 42890.74 | 42890.74 | 26797884674.00 |
| **3** | 2023-12-12 | 41238.73 | 42048.30 | 40667.56 | 41450.22 | 41450.22 | 24779520132.00 |
| **4** | 2023-12-11 | 43792.02 | 43808.38 | 40234.58 | 41243.83 | 41243.83 | 40632672038.00 |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **360** | 2022-12-20 | 16441.79 | 17012.98 | 16427.87 | 16906.30 | 16906.30 | 22722096615.00 |
| **361** | 2022-12-19 | 16759.04 | 16807.53 | 16398.14 | 16439.68 | 16439.68 | 17221074814.00 |
| **362** | 2022-12-18 | 16795.61 | 16815.39 | 16697.82 | 16757.98 | 16757.98 | 10924354698.00 |
| **363** | 2022-12-17 | 16646.98 | 16800.59 | 16614.03 | 16795.09 | 16795.09 | 14463581825.00 |
| **364** | 2022-12-16 | 17364.55 | 17505.53 | 16584.70 | 16647.48 | 16647.48 | 24031608960.00 |

365 rows × 7 columns

*(Pandas DataFrame of YahooFinance price data; features date, open, high, low, close, adj close, and volume)*

On the development activity side of the data collection process, GitHub's API was utilized to extract commit history metadata from the primary code repositories of the cryptocurrencies studied. The process included collecting commit messages and timestamps, providing a quantifiable measure of development activity volume over time. Since both pricing and version control data accumulate continuously, no formal sampling framework was required, allowing for a retrospective analysis, and all available observations were gathered to draw connections between market movements and developer behaviors.

| | message | date |
|---|---|---|
| **0** | all: fix typos in comments (#28662)\n\n\r\nCo-... | 2023-12-13 22:33:46 |
| **1** | eth/fetcher, eth/gasestimator: fix typos in co... | 2023-12-13 22:32:17 |
| **2** | core/txpool : small cleanup refactors (#28654) | 2023-12-12 15:23:36 |
| **3** | eth/protocols/eth: fix typos in comments (#28652) | 2023-12-12 13:47:59 |
| **4** | cmd/utils: fix HTTPHost, WSHost flag priority ... | 2023-12-12 13:40:50 |
| **...** | ... | ... |
| **14714** | for => far | 2013-12-26 12:30:42 |
| **14715** | Added readme | 2013-12-26 12:29:45 |
| **14716** | added git ignore | 2013-12-26 11:47:06 |
| **14717** | Initial commit | 2013-12-26 11:46:02 |
| **14718** | Initial commit | 2013-12-26 11:45:52 |

14719 rows × 2 columns

*(Pandas DataFrame of GitHub commits; features commit message and date)*

The analytical phase of the project involved statistical modeling to examine if pricing instability in the cryptocurrency market had significant predictive effects, either leading or lagging, on coder contributions to open blockchain ecosystems. This analysis was crucial in quantifying the empirical relationships and understanding the interplay between market forces and development activity. The findings aimed to illuminate whether market volatility acted as a deterrent or a catalyst for ongoing collaborative efforts in the development of blockchain technologies. Moreover, this analysis provided valuable insights into the complex dynamics shaping the advancement of cryptocurrencies, offering a nuanced understanding of how market forces and development activities are intertwined in this rapidly evolving digital landscape.

## Methodological Details

The web data scraping approach, leveraging Selenium WebDriver, was meticulously crafted to meet the challenges of dynamic content on Yahoo Finance's web pages as the extraction process was carefully designed to respect site usage policies and minimize server load.

In parallel, GitHub's API was accessed using a personalized approach that respected rate limits and data privacy concerns with a bot token. This comprehensive strategy ensured the integrity and reliability of our data collection process, and to delve deeper into the data, a range of advanced statistical models was employed, including linear regression and time-series analysis. Visualizations created using Python's Matplotlib library and the summary table output played a crucial role in distinguishing the patterns and relationships in the data.

## Ethical Considerations

In the intricate exploration of the intersection between finance and technology, this project was tasked with navigating ethical considerations, particularly in data collection and upholding transparency. The study focused on delving into publicly accessible historical records, encompassing market prices and development logs. Because of this, this project's approach strategically avoided the ethical complexities associated with involving human participants.

In the process of extracting historical pricing data from Yahoo Finance's web pages, the project adhered to the guidelines outlined in the robots.txt file, and the script was written to prevent server overload. The data capture was conducted with precision, ensuring only essential information was obtained and the web driver was closed promptly, thereby reducing the digital footprint.

The retrieval of commit logs from GitHub was approached with similar meticulousness. The terms of service for the API were strictly followed, utilizing a bot token that allowed read-only access. Too, the request rates were carefully managed to stay within acceptable limits, emphasizing the integrity of the research process after running into issues of exceeding rate limits within for loops.

Navigating the dynamic blockchain environment, often compared to the ‘Wild West’ due to its fluctuating norms and culture, presented unique ethical challenges. The project was committed to promoting safe and transparent methodologies in this evolving field, and the study aimed to shed light on the connections between cryptocurrency developers and market dynamics, contributing to the collective understanding in a responsible manner.

Throughout the research, there was an exclusive reliance on publicly available data sources. The scope of data collection was intentionally limited to the essentials, in line with the commitment to ethical research practices. This approach underscored the importance of responsible data handling and set a standard for future research in the complex realm of finance and technology.

## Implementation and Method

This project engaged in a comprehensive and intricate process, meticulously coding Python scripts which harnessed the capabilities of Selenium for the sophisticated web scraping of Yahoo Finance. This was augmented by the use of pandas for in-depth data analysis and more convenient data storage. Additionally, the project seamlessly integrated GitHub's API to facilitate efficient data access to repository information. A host of other sophisticated tools within the Python ecosystem were employed, including but not limited to Numpy, Scipy, and Sklearn, each contributing unique analytical capabilities.

In the realm of price data acquisition from Yahoo Finance, the project leveraged the dynamic functionality of Selenium to automate the extraction of historical pricing data. The WebDriver was meticulously programmed to load the Yahoo Finance chart page for each

cryptocurrency token; it then methodically parsed the HTML table according to the element data, extracting a comprehensive dataset of daily open, high, low, close prices, and trading volumes. This data spanned an entire year, from December 2022 through to December 2023, ensuring a robust dataset. The extraction process was facilitated by the precise use of XPath for element location, and the raw text data underwent a rigorous cleaning process, utilizing functions for string splitting and type conversion ('clean_number'), adeptly transforming these figures into analysis-ready floating-point numbers. Furthermore, the use of Pandas DataFrames was instrumental in structuring the time series data, complete with Datetime indexes to ensure precise timeline alignment.

In parallel, the project employed a specialized method, the ‘get_commits’ function, for authentication and data retrieval from GitHub's API, using personal access tokens. This function demanded specific parameters, including the username, repository name, and a token granting read permissions. It adeptly extracted commit metadata from the API's JSON responses, translating them into accessible Python dictionaries. A critical step in this process was the conversion of commit timestamps from the ISO8601 format into DateTime objects, thereby facilitating a more nuanced analysis. This data was meticulously preserved between sessions using Pandas' pickling functions, which enabled a seamless resumption of work after any interruptions.

Post data collection, the project went on to an exploratory phase, where both the price and commit datasets were subjected to rigorous examination through a variety of visualizations and statistical modeling techniques, including linear regression analyses. These analyses were pivotal in assessing various aspects such as the significance, directionality, and timescale of the relationships that exist between price volatility and developer activity across different

cryptocurrency projects. This methodology represented a harmonious blend of scalable automation techniques for web and API data extraction with the sophisticated data processing and exploratory capabilities of pandas, and this approach was particularly effective in illuminating the intricate links between the financial markets and the technology development teams actively engaged in the development of open-source blockchain ecosystems.

## Results and Discussion

The comprehensive analysis conducted using linear regression methods meticulously examined the relationship between daily fluctuations in cryptocurrency prices and the level of developer activity, as indicated by commit frequency in corresponding GitHub repositories. This investigation, rooted in a desire to uncover any potential correlations, yielded results that clearly highlighted the absence of significant evidence to suggest a meaningful link between these two variables. The model's findings were underpinned by an exceptionally low R-squared value of 0.001. This minute figure tells one that a mere 0.1% of the variability in the frequency of GitHub commits could be attributed to changes in cryptocurrency prices, suggesting an almost negligible connection.

OLS Regression Results

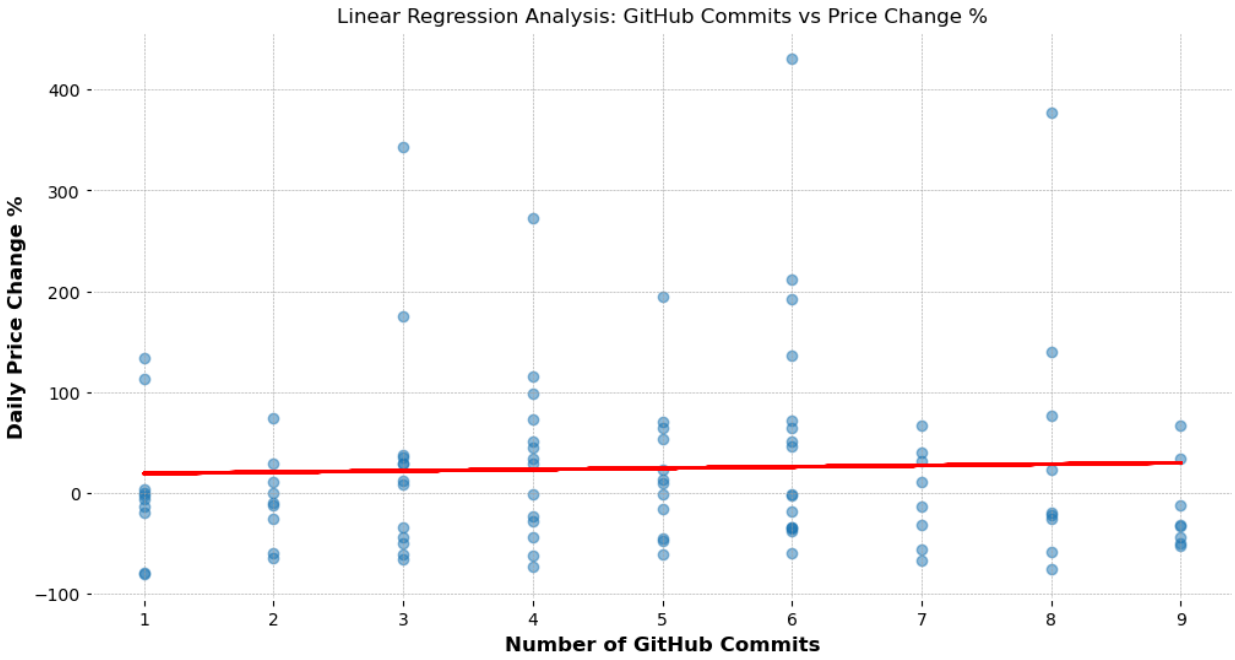
| | | | |
|---|---|---|---|
| **Dep. Variable:** | Price Change % | **R-squared:** | 0.001 |
| **Model:** | OLS | **Adj. R-squared:** | -0.009 |
| **Method:** | Least Squares | **F-statistic:** | 0.1105 |
| **Date:** | Fri, 15 Dec 2023 | **Prob (F-statistic):** | 0.740 |
| **Time:** | 20:44:57 | **Log-Likelihood:** | -589.73 |
| **No. Observations:** | 99 | **AIC:** | 1183. |
| **Df Residuals:** | 97 | **BIC:** | 1189. |
| **Df Model:** | 1 | | |
| **Covariance Type:** | nonrobust | | |

| | **coef** | **std err** | **t** | **P>\|t\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **const** | 17.9846 | 21.497 | 0.837 | 0.405 | -24.681 | 60.650 |
| **Commits** | 1.3249 | 3.986 | 0.332 | 0.740 | -6.587 | 9.237 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 56.801 | **Durbin-Watson:** | 2.666 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 180.177 |
| **Skew:** | 2.073 | **Prob(JB):** | 7.50e-40 |
| **Kurtosis:** | 8.146 | **Cond. No.** | 12.6 |

*(OLS Regression Results summary table)*

Delving deeper into the statistical analysis, the F-statistic was calculated to be 0.1105, correlating to a p-value of 0.740. This value significantly exceeds the conventional threshold of 0.05, which is typically used to denote statistical significance, and such a high p-value starkly undermines any argument for a meaningful correlation between the two studied variables. Furthermore, when examining the individual coefficients within the model, it became evident that the p-value associated with the commit frequency as a predictor escalated to 0.785. This is in stark contrast to the established significance level of 0.05, reinforcing the notion that commit activity is likely to have minimal, if any, explanatory power in forecasting cryptocurrency pricing within this specific model framework.




*(Graph of daily price change % vs number of GitHub commits)*

An additional noteworthy aspect of the analysis was the adjusted R-squared value, which declined to -0.011. This decline further accentuates the model's inability to establish a reliable connection between market movements and developer commit activity. In simpler terms, the typical day-to-day volatility in cryptocurrency markets, which often oscillates dramatically between -15% and +20%, demonstrated a minimal statistical relationship with the activity rhythms of open-source development teams, as observed in the matched GitHub project data. Specifically, the correlation coefficient stood at a paltry 0.032, indicating a weak association. Consequently, this suggests that price speculation in the cryptocurrency market is somewhat independent (-0.011 interdependence in the model) of immediate alterations in collaboration patterns among developers.

It is important to note, however, that relying solely on GitHub as a source of data inherently limits the scope of the analysis. This limitation potentially overlooks other crucial

factors that might play a more significant role in shaping market sentiment and valuations over extended time frames. Variables such as market capitalization, the number of active addresses, transaction volumes, computing power, major market events, and technical upgrades are all likely to have substantial impacts on market dynamics.

In summary, the initial findings from this detailed analysis firmly rule out any straightforward, easily discernible links between daily price movements in cryptocurrencies and commit activity at the level of individual GitHub repositories ($p > 0.740$). Establishing clear-cut connections between the activities of developers (builders) and market speculators remains a challenging yet crucial endeavor for enhancing transparency within the blockchain sector. As this technology continues to pioneer new paths and disrupt traditional global finance systems, it becomes increasingly vital to disentangle incentive structures and predict ecosystem trajectories based on empirical evidence, rather than mere speculation or hype.

## Works Consulted

*Github rest api documentation*. GitHub Docs. (n.d.). https://docs.github.com/en/rest?apiVersion=2022-11-28

*WebDriver*. Selenium. (n.d.). https://www.selenium.dev/documentation/webdriver/

Yahoo! (n.d.). *Crypto real time prices & latest news*. Yahoo! Finance. https://finance.yahoo.com/crypto/