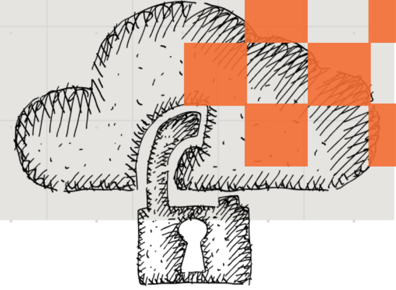
FCAT BLOCKCHAIN INCUBATOR



FCAT BLOCKCHAIN

Tracking Blockchain Health Part 2

A Quantitative Study of Ethereum Fundamentals.

By: Sophia Li, Joshua Chen, Di He, Aaron Gao, Ethan Atwood | April 2024

Introduction

Crypto enthusiasts live by a motto: “Code is Law”. In other words, code-based systems of record are held as superior to legacy financial systems because they are logic based, transparent and censorship resistant. This concept, and other memes, have led investors to pour multiple trillions into the crypto industry, and into native tokens that represent powerful and stable blockchains, like Bitcoin and Ethereum, as well as unproven blockchains with sometimes juvenile names, like Polkadot and Magic Internet Money. And while 95%[1] of blockchain startups ultimately fail, investors can try to predict the winners and losers using publicly available data. Because blockchains are built and operated in the public sphere, investors can track who is building and maintaining a network, who is using it, and who owns it. Investors ignore these fundamental risk parameters at their peril.

This research paper establishes a quantitative framework to test the relationship between blockchain network health metrics, otherwise known as blockchain fundamentals, and digital asset prices. It follows an earlier research paper, published in November 2023, which analyzed trends in blockchain fundamentals across leading Layer 1 blockchains. In this study, we analyze the relationship between Ethereum (the blockchain) and ETH (the native token) as a case study. With a better understanding of the blockchain metrics that impact native token prices, among other factors, investors might more appropriately focus their research efforts and generate a differentiated investment thesis.

The paper consists of seven sections.

1: Introduction
2: Summary of Findings
3: Blockchain Health Metric Descriptions
4: Correlation and Causation Analysis Methodology
5: Analytic Observations
6: Conclusion
7: Appendix

Summary of Findings

The relationships between blockchain health metrics and token prices are summarized in Table 1. All of the metrics show weak correlations with digital asset prices; Two metrics – Total Value Locked and Code-Commits, show mild causal relationships with prices, and warrant further investigation as potential leading indicators of price movement.

Table 1. Summary of Correlation and Causality Findings between Ethereum blockchain metrics and ETH price

Category	Metric	Correlation	Causality
Usage	Transaction Fees	Weak	Insignificant
	Number of Active Addresses	Weak	Insignificant
	Total Value Locked	Weak	Mild
Development	Active Developers	Weak	Insignificant
	Code-Commits	Weak	Mild
	Developer Experience	Weak	Insignificant
Decentralization	Validator Count	Weak	Insignificant
	Unlocked Supply Ratio	Weak	Insignificant

Note:

1. *Granger Causality is to be described in the Correlation and Causation Analysis Methodology section*
2. *Statistical significance of Granger Causality does not guarantee strong predictive power.*
3. *Results may be influenced by the historical data sampled for the study. This study uses daily volume weighted median prices of Ethereum.*

Blockchain Health Metric Descriptions

Usage

Transaction Fees

Blockchain Transaction Fees are paid to miners or validators for verifying transactions. Similar to convenience fees paid to payment processing corporations, like Visa and Mastercard, the fees serve as an incentive for miners or validators to process and confirm transactions. Transaction Fees are intended to contribute to the security and sustainability of a blockchain.

Number of Active Addresses

The Number of Active Addresses is defined as the number of unique wallet addresses engaging in transactions on a blockchain within the last 24 hours. The measure is often used to evaluate adoption and activity levels on a blockchain.

Total Value Locked

Total Valued Locked refers to the U.S. Dollar value of tokens held in liquidity pools, staking mechanisms, or otherwise controlled and custodied by a smart contract platform. The value can serve as a gauge of investors' confidence in a blockchain or in the opportunities built within that blockchain's ecosystem.

Development

Active Developers

Active Developers are measured as individual developers who have made code commits in the last 30 days to a blockchain's core repositories. This measure reflects activity in the code community, the robustness of the technology, and the level of continued innovation and investment in a project.

Code-Commits

Code-Commits are defined as changes submitted to a project's GitHub repository. Similar to the active developers measure, this metric tracks a blockchain's level of continuous innovation and maintenance. It is important to note, some blockchains have intentionally rigid codebases and that will suppress this measure relative to that of more open blockchains.

Developer Experience

Developer Experience is measured by the average number of years that core developers have contributed to a project. The measure reflects the degree of commitment of existing developers, as well as the addition and subtraction of new developers within a blockchain ecosystem.

Decentralization

Validator Count

The Validator Count is the number of wallets staking their assets to a blockchain and participating in blockchain security.

Unlocked Supply Ratio Market Share

The Unlocked Supply Ratio Market Share represents the proportion of a specific cryptocurrency's unlocked supply ratio relative to the sum of the unlocked supply ratios of other major cryptocurrencies in the market. The ratio can be used to measure how susceptible a blockchain is to inflation and market manipulation.

Correlation and Causation Analysis Methodology

Momentum

This research quantifies momentum as the percentage differences between current values and trailing averages of a blockchain metric:

$$Momentum = \frac{Current\ Value - Trailing\ N\text{-}Day\ Average}{Trailing\ N\text{-}Day\ Average}.$$

Comparing a current value to its multi-day historical average is preferred to a single-day comparison, because the resulting momentum measure will be more robust to outliers and better suited for subsequent correlation and causality analyses. In the causality analysis, a 30-day calculation window was chosen to balance the tradeoff between ensuring a sufficient sample size and reducing noise in the data.

Pearson Correlation Coefficient

The Pearson Correlation Coefficient is calculated to evaluate the degree to which changes in a health metric and changes in price are related. Given a pair of variables (X, Y), the Pearson correlation coefficient is calculated as:

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y},$$

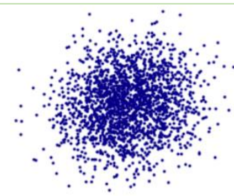
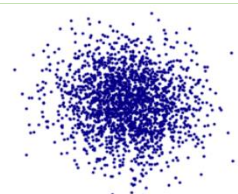
where

- $cov(X,Y)$ is the covariance (joint variability) between X and Y,
- σ_X is the standard deviation of X,
- and σ_Y is the standard deviation of Y.

Pearson Correlation Coefficients range between -1 and +1. A larger absolute value of the correlation coefficient indicates a stronger relationship between two variables. Table 2 below summarizes a generally accepted guideline for interpreting the measure.

Table 2. Guideline for Interpreting Pearson Correlation Coefficients

Range	Interpretation	Illustration
0	No linear relationship	
+1	Perfect positive linear relationship	
-1	Perfect negative linear relationship	
(0.7, +1)	Strong positive linear relationship	

Range	Interpretation	Illustration
(-1, -0.7)	Strong negative linear relationship	
(0.3, 0.7]	Moderate positive linear relationship	
[-0.7, -0.3)	Moderate negative linear relationship	
(0, 0.3]	Weak positive linear relationship	
[-0.3, 0)	Weak negative linear relationship	

Granger Causality Test

While Pearson correlation analysis is an effective measure of the degree of co-movement between two variables, it does not detect if the movement of one variable leads to that of the other. The Granger Causality Test is designed to determine causal relationships between variables. There is an assumption within this test that if one variable both precedes and provides useful information in forecasting another, then there is a causal effect between them. In other words, if lagged values of a variable X provide information on current values of another variable Y, then statisticians describe this effect as "X 'Granger-causes' Y." It is worth noting that this is an isolated measure, and the observed relationship may be driven by other extraneous factors in the environment.

In our study, this research investigated the Granger causality between the momentum of blockchain metrics and the momentum of Ethereum price. Because daily data has noises and doesn't have Granger Causality in our test. Monthly data can better capture trend and reduce noises. Therefore, this paper used monthly data instead of daily data.

A prerequisite for performing the Granger Causality test is that the data need to be stationary. This paper transformed the non-stationary blockchain metrics and Ethereum price to stationary data by applying the momentum formula, and the Augmented Dickey-Fuller (ADF) test was used to confirm that the momentum data is stationary (Refer to Appendix).

To interpret the output of the Granger Causality Test applied in this research, we need to define output measures to be shared. The research objective is to show the statistical significance and strength of the links between the health of a blockchain and token price changes.

- The p-value of the Granger Causality Test describes statistical significance. If the p-value is lower than 0.05, one can establish statistical significance of causality at 95% confidence, the widely applied level of significance in statistics.
- Adjusted R^2 from the research linear regressions indicates the explanatory power of blockchain health on token price movements. The calculation can range from an unbounded negative value to +1. The higher the adjusted R^2, the stronger the explanatory power.
- The Granger Causality Test contrasts two multivariate regression models in producing adjusted R^2.
    - A "restricted" model only used past values of token prices at varying lags at a 1-month interval.
    - An "unrestricted" model used past values of token prices in conjunction with past values of blockchain health.

  The strength of explanatory power is measured by the lift of the unrestricted model's adjusted R^2 over that of the restricted model.

A supplemental measure may need to be used to further validate and explain causal relationships. Though the Granger Causality Test is based on multivariate regressions that uses all the time lags, the research performed a univariate regression to further validate the causality of the statistically significant time lag to reduce the chance of model overfitting.

Analytic Observations

Usage

Transaction Fees

For transaction fees, the covariation with price movements remains weakly negative as illustrated in Figures 1 and 2.

The Granger Causality Test, as shown in Table 3, does not indicate a causal relationship between transaction fees of the Ethereum blockchain and the native token (ETH) price (demonstrated by the statistically insignificant p-value). Moreover, the adjusted R^2 is low, also indicating weak explanatory power of transaction fees.

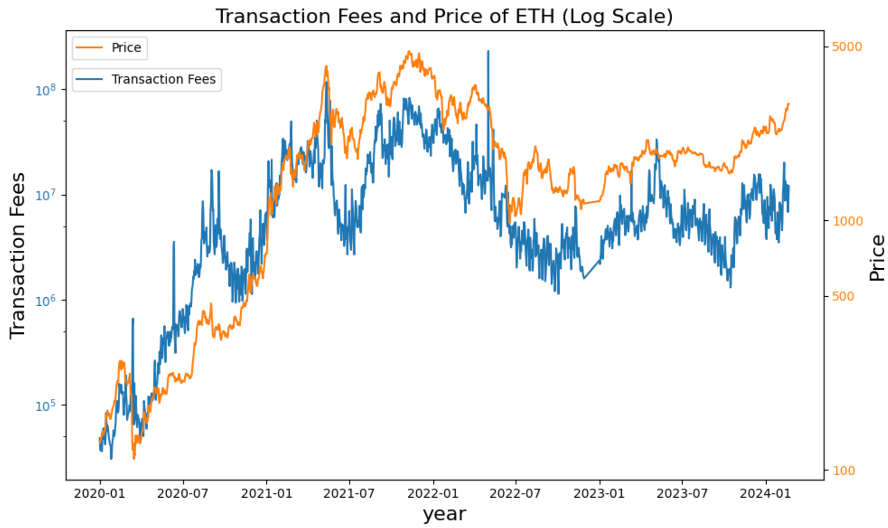


Figure 1. ETH Transaction Fee and Price

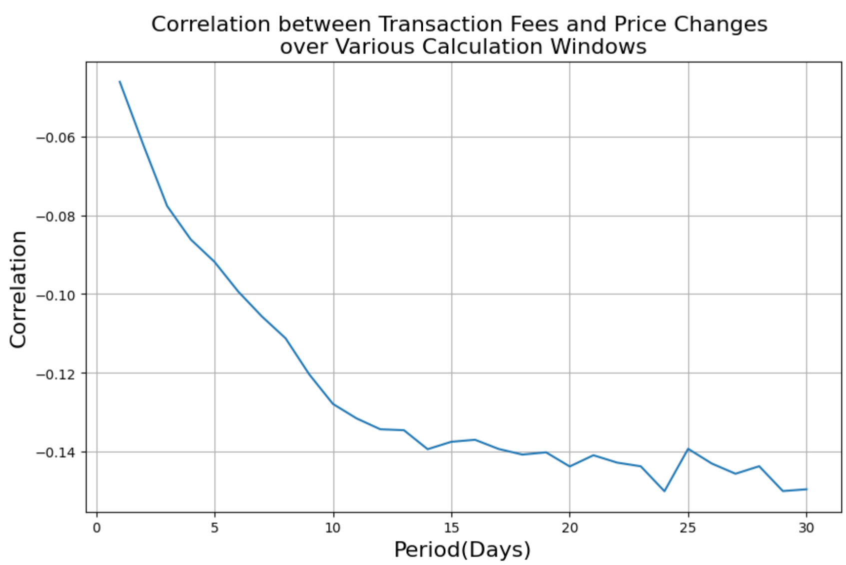


Figure 2. Correlation between ETH Transaction Fee and Price Momentums over Various Calculation Windows

Table 3. Granger Causality Test Result – Transaction Fees

	1 month lag	**1~2 month lags**
P-value (Granger Causality Test)	0.0642	0.0548
adjusted R^2 (Restricted Model)	-0.0195	-0.0254
adjusted R^2 (Unrestricted Model)	0.0273	0.0601

Number of Active Addresses

The covariation between Ethereum's number of active addresses and ETH price is consistently weakly negative.

The p-value of the Granger Causality Test is consistently insignificant, and the adjusted R^2 is consistently low, indicating no causal effect and limited explanatory power.

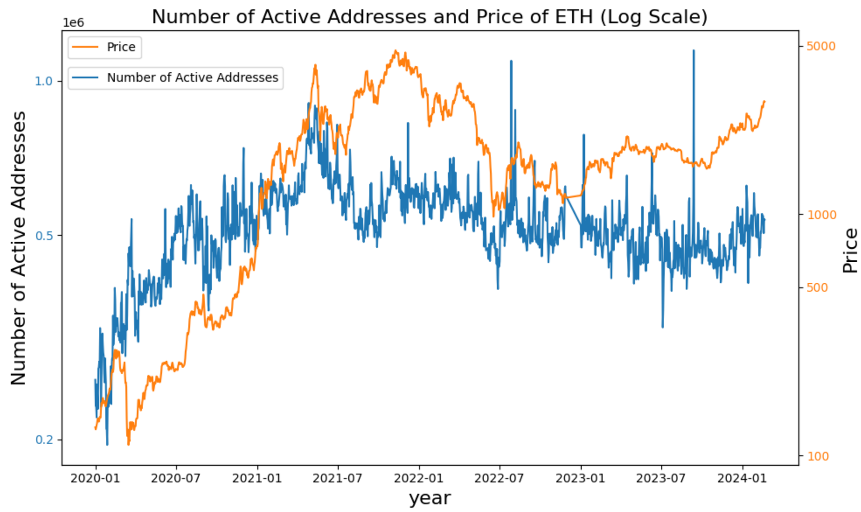


Figure 4. Number of Active Addresses and Price of ETH

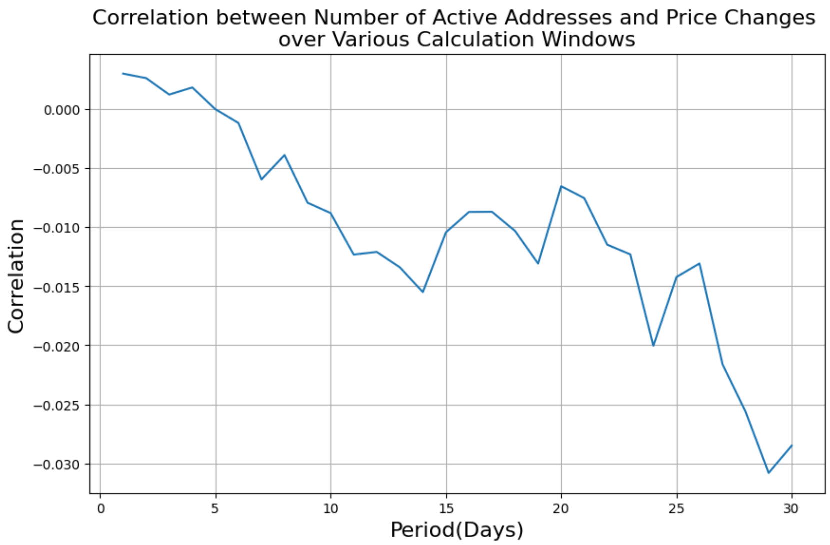


Figure 5. Correlation between Number of Active Address and Price Momentums for ETH over Various Calculation Windows

	1 month lag	**1~2 month lags**
P-value (Granger Causality Test)	0.7367	0.6703
adjusted R^2 (Restricted Model)	-0.0195	-0.0254
adjusted R^2 (Unrestricted Model)	-0.0397	-0.0563

Figure 6. Granger Causality Test Result – Number of Active Addresses

Total Value Locked

Total Value Locked and price movements are weakly negatively correlated as can be seen in Figures 7 and 8.

There exists a mild causal relationship between total value locked and token price momentum. The p-value from the Granger Causality Test is statistically significant. Moreover, when total value locked lagged by 1 month is tested in isolation, the p-value is also statistically significant, confirming a causal impact. However, this causal relationship should not be interpreted as having predicative power, as suggested by a low adjusted R^2.

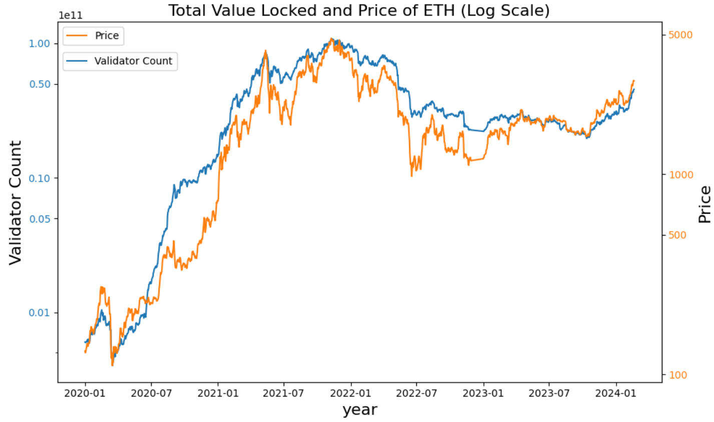


Figure 7. Total Value Locked and Price of ETH

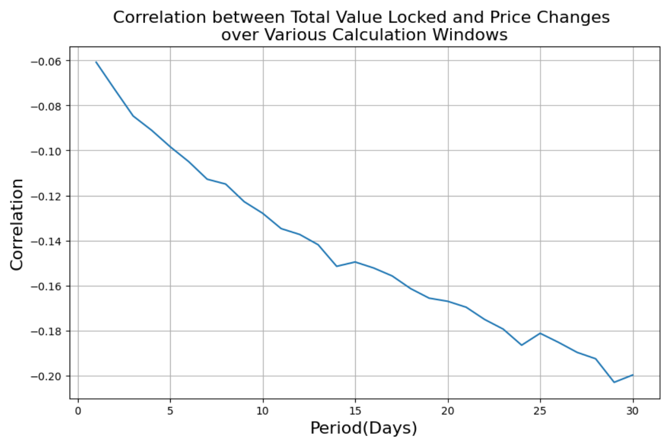


Figure 8. Correlation between Total Value Locked and Price Momentums for ETH over Various Calculation Windows

Table 3. Granger Causality Test Result – Total Value Locked

	1 month lag	**1~2 month lags**
P-value (Granger Causality Test)	0.0235	0.0113
adjusted R^2 (Restricted Model)	-0.0195	-0.0254
adjusted R^2 (Unrestricted Model)	0.0586	0.0978

Development

Active Developers

Active Developers and price movements are weakly negatively correlated as can be seen in Figures 10 and 11.

No causal effect can be confirmed between active developers and token prices. The p-value is consistently statistically insignificant, and the adjusted R^2 is low.

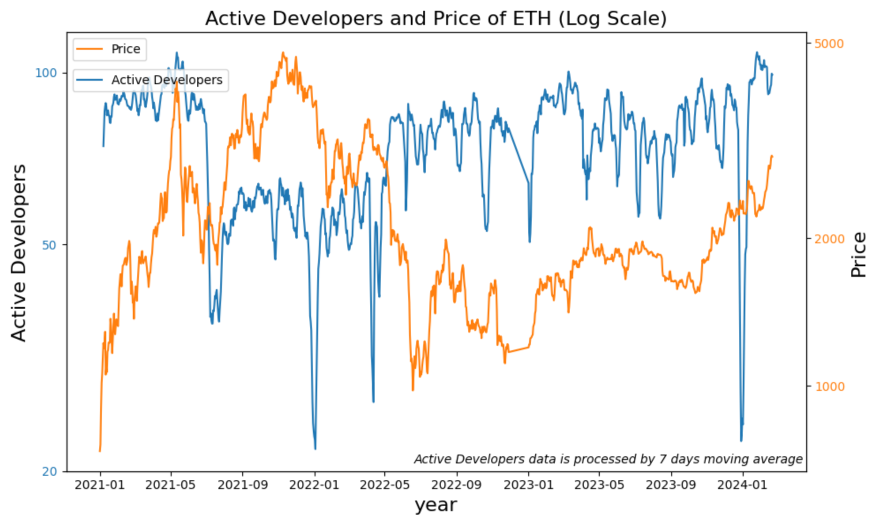


Figure 10. Active Developers and Price of ETH

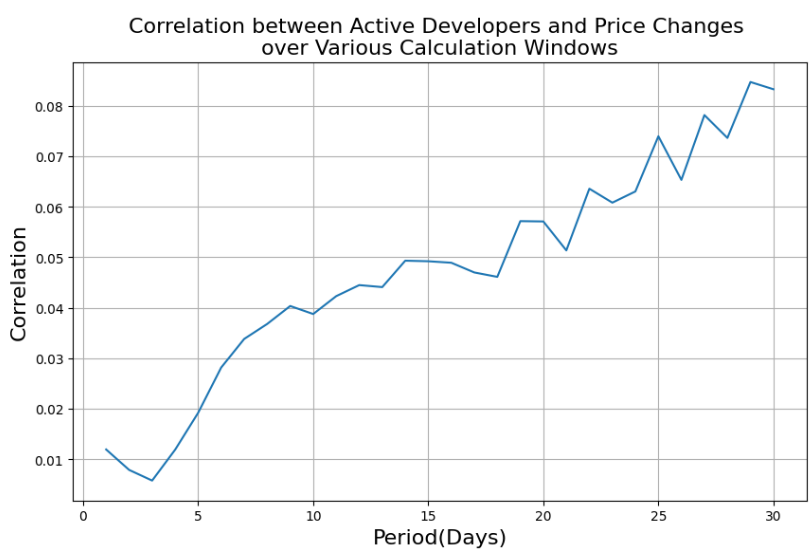


Figure 11. Correlation between Active Developers and Price Momentums for ETH over Various Calculation Windows

Table 12. Granger Causality Test Result – Active Developers

	1 month lag
P-value (Granger Causality Test)	0.2444
adjusted R^2 (Restricted Model)	0.0268
adjusted R^2 (Unrestricted Model)	0.0337

Code-Commits

Code-commits are found to have weakly positive correlations with digital asset prices measured over different calculation windows, as Figures 13 and 14 illustrate.

Although the Pearson Correlation between code-commits and price momentums is low, the Granger Causality Test indicates a mild causal effect between them. The p-value is statistically significant at a lag of 1 month, both when the lagged values of code-commits are tested in conjunction and in isolation. However, this causal relationship should not be interpreted as having predicative power, as suggested by a low adjusted R^2.

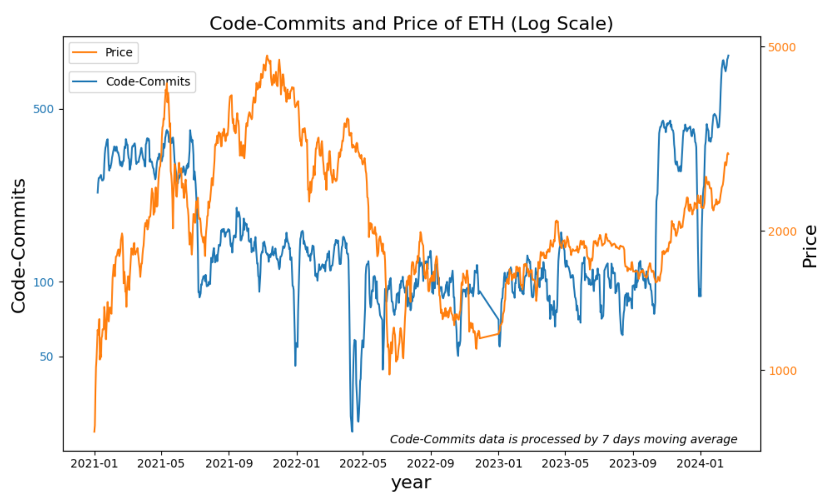


Figure 13. Code-Commits and Price of ETH

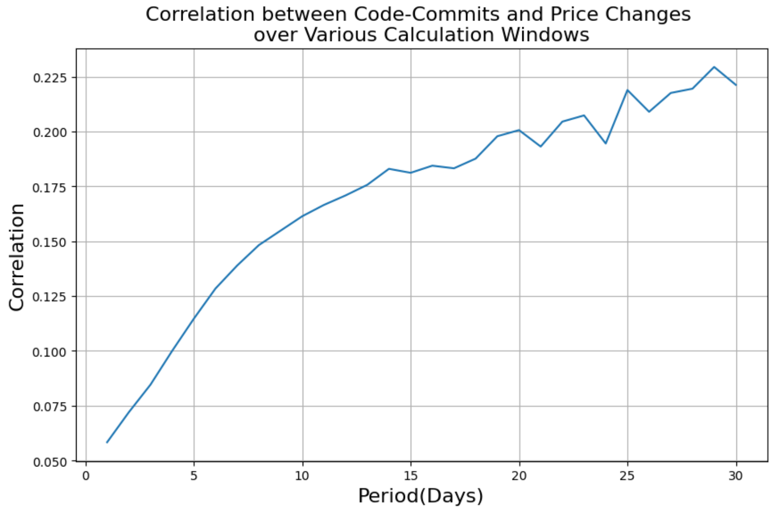


Figure 14. Correlation between Code-Commits and Price Momentums for ETH over Various Calculation Windows

Table 15. Granger Causality Test Result – Code-Commits

	1 month lag
P-value (Granger Causality Test)	0.0458
adjusted R^2 (Restricted Model)	-0.004
adjusted R^2 (Unrestricted Model)	0.0706

Developer Experience

Developer experience exhibits weakly negative correlations with price movements, as Figures 16 and 17 illustrate.

The p-value of the Granger Causality Test is consistently insignificant, and the adjusted R^2 is consistently low, indicating no causal effect and limited explanatory power.

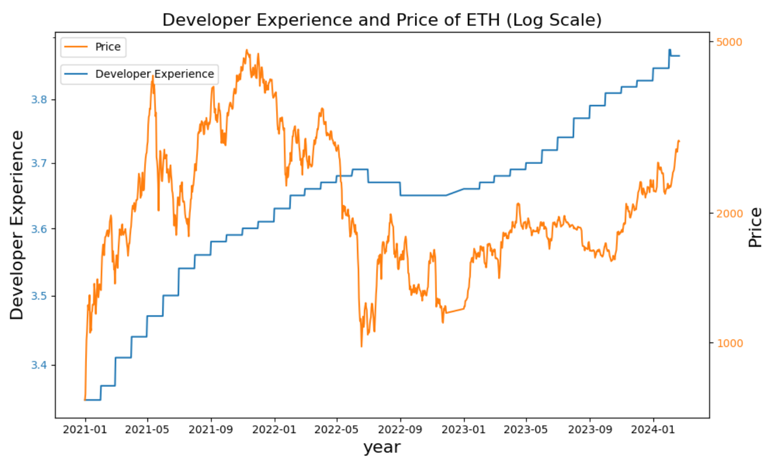


Figure 16. Developer Experience and Price of ETH

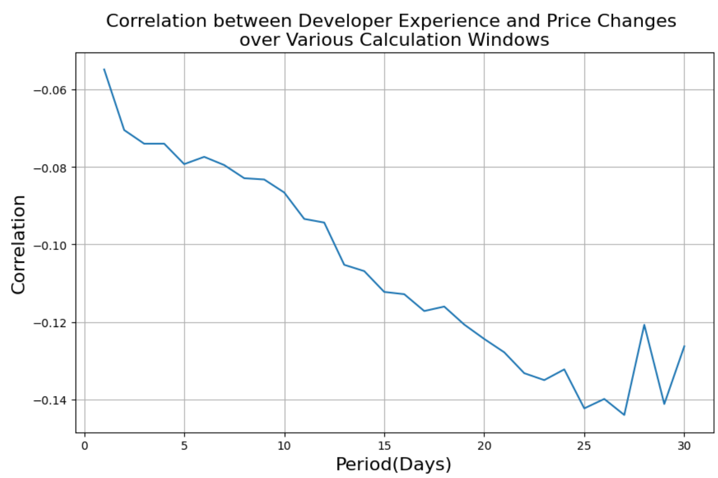


Figure 17. Correlation between Developer Experience and Price Changes for ETH over Various Calculation Windows

Table 18. Granger Causality Test Result – Developer Experience

	1 month lag
P-value (Granger Causality Test)	0.2166
adjusted R^2 (Restricted Model)	0.0268
adjusted R^2 (Unrestricted Model)	0.0383

Decentralization

Validator Count

Validator count and price movements are weakly positively correlated as can be seen in Figures 19 and 20.

No causal effect can be confirmed between validator count and token prices. The p-value is consistently statistically insignificant, and the adjusted R^2 is low.

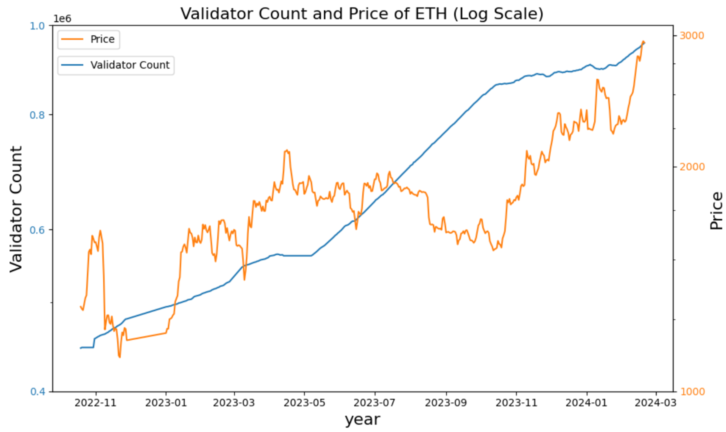


Figure 19. Validator Count and Price of ETH

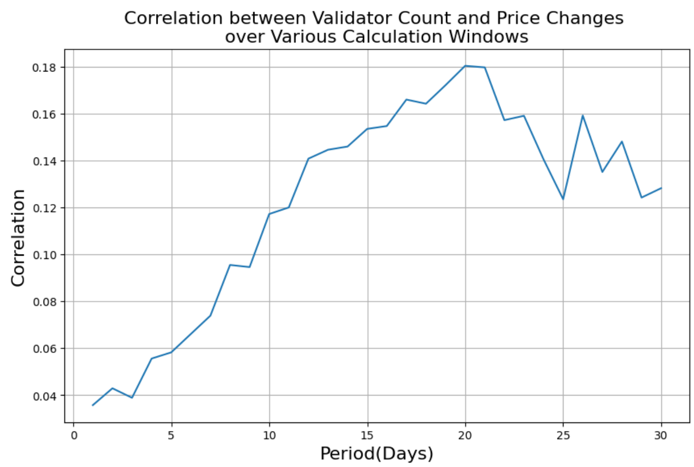


Figure 20. Correlation between Validator Count and Price Changes for ETH over Various Calculation Windows

Table 21. Granger Causality Test Result – Validator Count

	1 month lag
P-value (Granger Causality Test)	0.0737
adjusted R^2 (Restricted Model)	-0.0586
adjusted R^2 (Unrestricted Model)	0.0599

Unlocked Supply Ratio

There exists a weakly positive correlation between the unlocked supply ratio market share and price movements over different calculation windows.

With a statistically insignificant p-value, we conclude that there is no causal relationship between unlocked supply ratio market share and price changes.

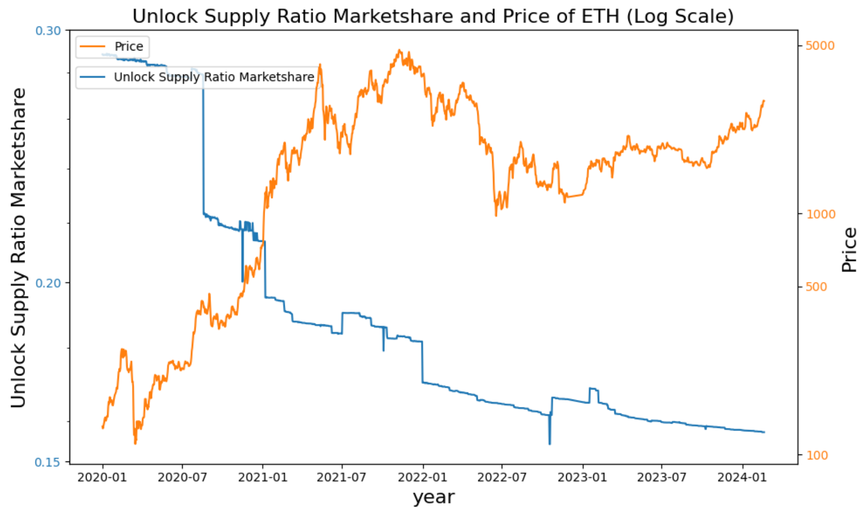


Figure 22. Unlocked Supply Ratio and Price of ETH

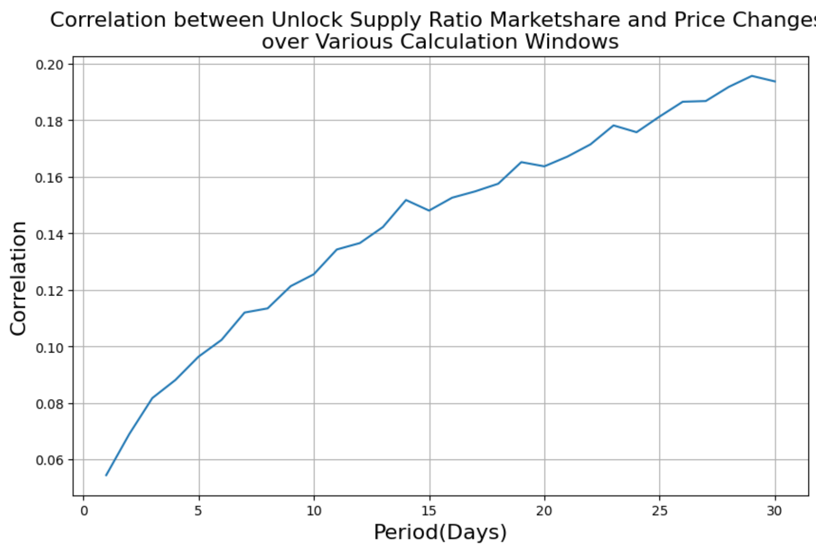


Figure 23. Correlation between Unlocked Supply Ratio Market Share and Price Changes for ETH over Various Calculation Windows

Table 24. Granger Causality Test Result – Unlocked Supply Ratio Market Share

	1 month lag	**1-2 month lags**
P-value (Granger Causality Test)	0.6423	0.8795
adjusted R^2 (Restricted Model)	-0.0195	-0.0254
adjusted R^2 (Unrestricted Model)	-0.0375	-0.0684

Conclusion

Across all examined metrics, Pearson Correlation analysis yielded weak and insignificant correlations with Ethereum price. By leveraging the Granger Causality test and additional univariate regression, this study also found mostly insignificant causal relationships between blockchain metrics and Ethereum price. However, Total Value Locked and Code-Commits were found to have statistically significant causal relationships.

While this study uncovered statistically causal relationships in isolation, there are multiple factors not considered in the analysis. The price of any digital asset is influenced by a complex web of market forces and investor sentiment, among other factors. Are investments in dApps built on top of a blockchain, or spikes in open-source developer activity, leading indicators of price appreciation? Are declines in dApp usage and developer activity signals that a blockchain is failing? We are investigating these questions and more at SherlockAnalytics. Please reach out to contactsherlock@fmr.com for more information.

Appendix

Monthly Momentum Data Stationary Test

Stationarity is a crucial property of time series data, where the statistical characteristics remain constant over time. In the context of Granger Causality analysis, ensuring the stationarity of the data is essential to avoid spurious relationships and misleading results. The Augmented Dickey-Fuller (ADF) test is a widely used statistical method to determine whether a time series is stationary or contains a unit root. The ADF test assesses the null hypothesis that a unit root is present in the data, and rejecting this hypothesis indicates that the series is stationary.

Table 11. ADF Stationary test result

Metric	Monthly Data Size	P-value
Price	49	4.115e-10
Transaction Fees	49	0.009136
Number of Active Addresses	49	0.0002643
Total Value Locked	49	0.001238
Active Developers	36	5.6569e-12
Code-Commits	36	0.0009748
Developer Experience	36	0.03009
Validator Count	15	0.01707
Unlocked Supply Ratio	49	0.0006453

Recommended Maximum Lag in the Granger Causality Test

Excessively long lags can lead to model overfitting. To mitigate this issue, we determined maximum time lags based on the rule of thumb that at least 10 observations are required for each independent variable in regression analysis (Harrell, 2001). In the context of Granger Causality testing, each additional lag introduces two variables to an unrestricted model. Consequently, 20 observations were necessary for each lag included in the analysis.

Table 12. Recommended Maximum Lag in the Granger Causality Test

Category	Metric	Start Date	End Date	Data Size	Monthly Data Size	Max Lag
Usage	Transaction Fees	2020-01-01	2024-02-20	1474	49	2
	Number of Active Addresses	2020-01-01	2024-02-20	1474	49	2
	Total Value Locked	2020-01-01	2024-02-20	1474	49	2
Development	Active Developers	2021-01-01	2024-02-20	1108	36	1
	Code-Commits	2021-01-01	2024-02-20	1108	36	1
	Developer Experience	2021-01-01	2024-02-20	1108	36	1
Decentralization	Validator Count	2022-01-01	2024-02-20	456	15	1
	Unlocked Supply Ratio	2020-01-01	2024-02-20	1474	49	2

References

Harrell, F.E. (2001) Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer-Verlag, New York. http://dx.doi.org/10.1007/978-1-4757-3462-1

Shawgador, Jinia. "Why 95% of Blockchain Startups Fail." Techopedia.com, 23 Nov. 2023, www.techopedia.com/why-95-of-blockchain-startups-fail#:~:text=Even%20with%20a%20high%20failure,1%2C200%20billion%20USD%20by%202030.