Effect of Crypto-currencies Market Price and User Behavior into DeFi Lending Protocols: Final report

Ulysse Kazmierczak
Tsitohaina Ravelomanana
Bilal Benhana

May 16, 2024

Contents

1 Introduction

The traditional banking system, born out of the need for a trusted intermediary to validate currency legitimacy, has long dominated global finance. However, its susceptibility to failures, as evidenced by historical market crashes and economic crises, has prompted a quest for alternative solutions. In the aftermath of the 2008 subprime crisis, Satoshi Nakamoto introduced Bitcoin, a digital currency operated through a peer-to-peer network governed by mathematical algorithms [3].

The subsequent evolution of Ethereum, with its smart contract capabilities, opened up a realm of possibilities beyond mere peer-to-peer transactions. Smart contracts paved the way for decentralized finance (DeFi) protocols, announcing a new era of financial innovation [6]. These protocols, epitomized by platforms such as Compound V2 [2], reimagine financial services using blockchain technology to eliminate traditional intermediaries, such as banks and brokerage houses.

At the forefront of this financial revolution lies decentralized lending, a cornerstone of DeFi applications. Compound V2 is a shining example, empowering users to seamlessly participate in cryptocurrency lending and borrowing at variable interest rates, all executed transparently on the Ethereum blockchain.

This paper explores user behavior dynamics within Compound V2, delving into the intricacies of one of DeFi's most prominent lending protocols. Building upon previous research [7], we aim to construct a robust statistical model elucidating how users engage with Compound V2's lending pools, particularly in response to market fluctuations in cryptocurrency prices. Through meticulous analysis of balance sheets and aggregated user activities, this study endeavors to unearth insights into how users navigate the high volatility inherent in cryptocurrency markets and the consequential impact of their decisions on asset prices.

By shedding light on the nuanced interactions within the DeFi ecosystem, this research helps develop our understanding of decentralized finance. It offers valuable insights into its broader implications for traditional financial markets. As the landscape continues to evolve, a comprehensive grasp of DeFi dynamics becomes indispensable for shaping the future of the global economic system.

1.1 Background on the Compound Protocol

Before delving into our research, it's crucial to grasp the inner workings of Compound. A simplified overview of Compound is depicted in Figure 1.

Compound operates as a lending protocol with three key participants: suppliers, borrowers, and liquidators. Suppliers contribute liquidity to the protocol by depositing their crypto assets. They can withdraw their liquidity whenever, earning yields based on duration and the amount supplied. Conversely, borrowers can deposit assets as collateral and borrow different assets within the protocol. They can repay their borrowed assets with interest and retrieve their collateral. Liquidators, though not central to our study, serve as protocol authorities and liquidate borrowing positions lacking sufficient collateral.

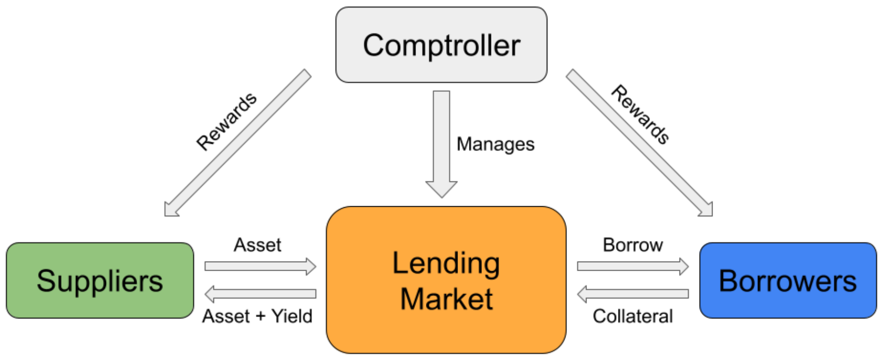


Figure 1: Compound Visualizer (Simplified Version)

To incentivize participation, Compound introduced the Comptroller, an entity that rewards users with COMP tokens based on their borrowing and supplying activities, contingent upon the protocol's state. This behavior is inconsistent with rational borrowing practices due to higher borrowing interest rates than supply interest rates and underscores the fact that DeFi users engage in yield-seeking strategies.

1.2 Interest Rates in Decentralized Finance

Besides, decentralized finance cannot be tackled the same way as traditional finance. Interest rates on DeFi lending and deposit platforms exhibit notable volatility, often diverging significantly from traditional finance benchmarks. On platforms like Compound, interest rates are determined by a piecewise linear function of the utilization ratio, a measure reflecting the ratio of borrowed assets to supplied assets within the protocol [2].

As Barthélemy et al. reported [1], Figure 2 shows the typical intraday fee for borrowing a stablecoin on AAVE and Compound hovered around 5% in January 2022, while the Fed funds rate remained at zero. This discrepancy underscores the distinctive nature of interest rate dynamics within DeFi, driven by platform-specific supply and demand conditions rather than centralized policy decisions.

Understanding the drivers behind changes in the utilization ratio is crucial for comprehending user engagement and optimizing lending and borrowing strategies within the DeFi ecosystem.

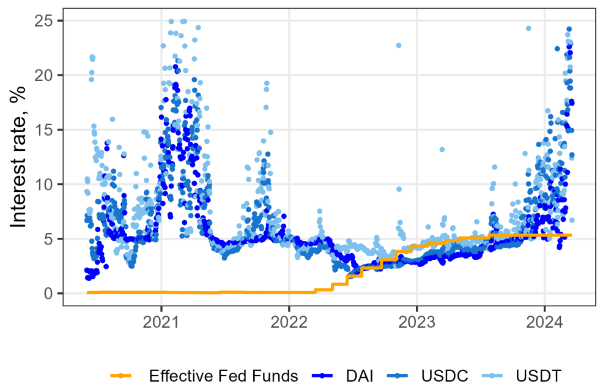


Figure 2: Interest rates on decentralized finance platforms (Source: [1])

1.3 Research Questions

Our study will delve deeper into several mechanisms within the Compound lending protocol:

- **Crypto-Currency Features and User Behavior:** First, we aim to explore the relationships between the features of various cryptocurrencies (focusing on COMP, BTC, ETH, DAI, USDT, and USDC) and user behavior on Compound V2 by examining the utilization ratio of their respective pool.

- **Impact of Ethereum Price Fluctuations:** Given that all assets supported within the protocol are Ethereum-backed cryptocurrencies, we seek to analyze the impact of Ethereum price fluctuations on borrowing activities within Compound.

- **Liquidity Mining Practices:** Lastly, we investigate the peculiar practice of liquidity mining, where users borrow against the same asset as collateral. Although this behavior appears counterintuitive from the protocol's perspective, it is driven by the incentive structure established by the Comptroller, as borrowers aim to acquire COMP tokens.

We aim to comprehensively understand user behaviors and their implications within the Compound ecosystem by addressing these questions.

2 Related Works

Our study builds upon the insights provided by Tovanich et al., "Contagion in Decentralized Lending Protocols: A Case Study of Compound" [7]. This study detailed the construction of Compound's liquidity pool balance sheets and identified instances of liquidity mining misuse. Leveraging the balance sheets they developed, we want to better understand the behaviors of Compound users, complementing our analysis with daily data sourced from the Ethereum ledger.

Saengchote [5] provides a foundational analysis of user behavior on Compound. The paper investigates the impact of interest rates on deposit demand, the relationship between market volatility and stablecoin deposit demand, and the influence on stablecoin loan demand. Despite these efforts, the study concludes that consistent patterns across all pools are elusive, primarily due to the confounding effects of yield farming incentives and the inherent differences between various coins. We build on Saengchote's work by aiming to isolate genuine borrowing activity from self-borrowing and yield farming behaviors. By doing so, we seek a precise understanding of user behavior and the true dynamics of liquidity within the Compound protocol.

Additionally, to explore the phenomenon of liquidity mining, we refer to the research conducted by Park and Stinner [4]. Their work sheds light on the efficacy of liquidity mining programs, highlighting how liquidity reward incentives significantly deposits and borrowing activities within DeFi protocols. Notably, they found that yield-seeking participants, particularly attracted by lucrative incentives such as those offered by stablecoins, played a substantial role in driving deposits and borrowings during the peak of DeFi enthusiasm. We want to validate these findings with our dataset and examine how various attributes influence liquidity mining dynamics.

By synthesizing insights from previous research with our empirical analysis, we aim to contribute to a deeper understanding of decentralized lending protocols and their implications for the broader DeFi ecosystem.

3 Data Description

To ensure the robustness of our study, we rely on accurate and comprehensive data regarding the activities of Compound users. We utilize the balance sheets from Tovanich et al. [7], which provide a detailed overview of Compound's financial network.

3.1 Pool and Market Features

To investigate coins market features and user behavior interplay, we categorize features into pool features and market features. The panel of features is described below:

Table 1: Panel of features used to investigate user behavior interplay.

Variable	**Description**
Coin Market Price	The market price of the cryptocurrency in US Dollars. It serves as a fundamental metric for assessing its value and market performance.
Daily Return	Daily Return measures the percentage change in the cryptocurrency's market price compared to yesterday's price. It could be considered as a proxy for market volatility.
Total Supply	The total supply of the cryptocurrency available within the liquidity pool indicates the quantity of the cryptocurrency deposited in the lending pool from users.
Total Borrow	The total Borrow of the cryptocurrency available within the liquidity pool indicates the quantity of the cryptocurrency borrowed in the lending pool from users.
Collateral Factor	The collateral factor indicates the proportion of deposited coins that can be used as collateral to borrow from the lending pool. The value ranges from 0 to 1.
Cash	Cash refers to the amount of liquid funds available within the lending pool that can be utilized for lending purposes or to meet withdrawal requests from users.
Reserves	Reserves represent a portion of assets held aside within the lending pool to cover potential losses or mitigate risks associated with borrower defaults.
Utilization Ratio 1	The utilization ratio provides insights into the demand for borrowing relative to the available supply in the pool. A high utilization ratio indicates strong borrowing demand and leads to higher interest rates.
Borrow Rate 2	The borrowing rate of a cryptocurrency is the rate at which a user can borrow cryptocurrency. It is a piecewise linear function of the utilization ratio.

$$U_{ratio} = \frac{\text{Total Borrows}}{\text{Total Borrows} + \text{Cash} - \text{Reserves}} \tag{1}$$

$$\text{borrow Rate} = \begin{cases} r_0 + U_{ratio} \times m & \text{if } U_{ratio} \leq k \\ r_0 + k \times m + (U_{ratio} - k) \times n & \text{if } U_{ratio} > k \end{cases} \tag{2}$$

where: r_0 is the base rate; m is the slope of the borrow rate before the kink; n is the slope of the borrow rate after the kink; k is the kink point.

These features are the characteristics of lending pools, which are useful for understanding the mechanisms and users' collective behaviors within the Compound lending protocol.

3.2 Liquidity Mining Activity Features

To assess the liquidity mining activities, we extract data on liquidity mining activities from Tovanich et al.'s daily balance sheets dataset [7]. We calculate the share of transactions involving a user borrowing against the same collateral.

We also collect data about the COMP token market from external sources to analyze the following factors that could influence liquidity mining activities:

- **Total Locked Value:** We obtain the total locked value (TVL) of Compound protocol from DefiLlama[1], a DeFi data aggregator. TVL is a proxy for Compound's usage and adoption within the DeFi ecosystem.
- **COMP Token Price:** We rely on CoinGecko[2] for data on the COMP token price, enabling us to analyze the relationship between token price movements and liquidity mining activities within Compound.
- **Daily Distribution of COMP Tokens:** CoinMarketCap[3] provides data on the total number of COMP tokens available in the protocol. We use this information to compute the daily distribution of COMP tokens.

We hope these features drive the liquidity mining activity and will help enhance our understanding of this peculiar practice.

[1] https://defillama.com/
[2] https://www.coingecko.com/
[3] https://coinmarketcap.com/

4 Interplay of cryptocurrency and pool features

4.1 Data Visualization

To begin our analysis, we started by visualizing the price and utilization ratio trends to identify characteristic behavior. Figure 3 illustrates the evolution of prices with their respective utilization ratios (in USD) for volatile assets (ETH, BTC). Graphically, the prices of ETH and BTC appear to be positively correlated. However, their utilization ratios do not exhibit a clear correlation. This observation raises two possibilities: either the market price is not a suitable variable to explain the utilization ratio or a more in-depth statistical analysis is required to uncover potential relationships.

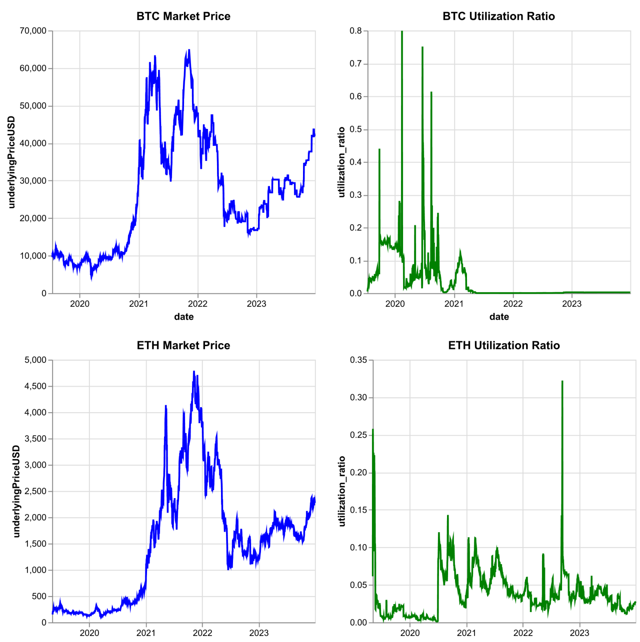


Figure 3: Market Price and Utilization Ratio (Volatile Coins)

For stablecoins which are pegged to 1$ all time long by definition, their utilization ratio seems to show a clearer correlation this time as we can see in Figure 4:

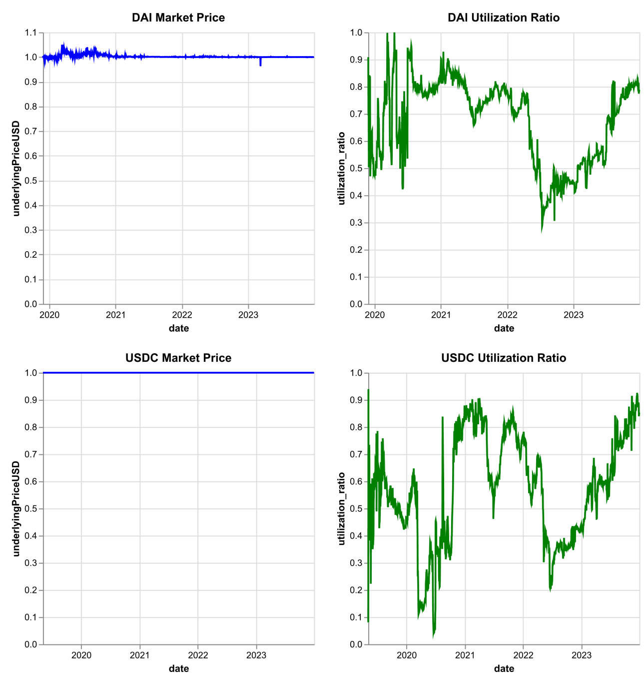


Figure 4: Market Price and Utilization Ratio (Stable Coins)

In contrast, Figure 5 presents a graphical representation of the utilization ratios of both coins along with the borrowing interest rates. This visualization highlights a functional dependence, linear piecewise, between these two features as described in the Compound V2 documentation [4].

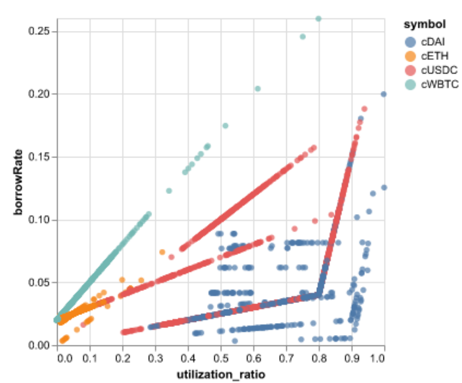


Figure 5: Utilization Ratio and Borrowing Interest Rates

[4]https://docs.compound.finance/interest-rates/

4.2 Time Series Assumptions for Granger Causality

We employ Granger causality testing to examine the lead/lag autocorrelation between two variables. It assesses whether changes in one variable precede changes in another, indicating a potential causal relationship among the following key features: collateral factor, total supply, market price, utilization ratio, and daily returns for a given coin. Before conducting the Granger causality tests, we checked the model's assumptions, such as stationarity, to ensure the validity of the results.

To prepare the time series data for Granger causality testing, data cleaning steps are performed. First, we normalize the time series by subtracting the mean of each series from every data point and then dividing by the standard deviation. Following, the trend is removed from each time series. This is achieved by taking the difference between consecutive data points (first difference operation).

To test the stationarity of the time series, we perform an Augmented Dickey-Fuller test. This test checks for the presence of a unit root in a time series, which indicates non-stationarity. If the test rejects the null hypothesis of a unit root, it suggests stationarity, validating subsequent analyses and predictions. In the algorithms, the Augmented Dickey-Fuller (ADF) test is employed as an initial step to assess the stationarity of a time series. It begins by testing whether the time series is stationary with a significance level typically set at 5%.

If the test indicates that the time series is not stationary, we perform the data-cleaning process specified above to enhance stationarity. This approach optimizes complexity by prioritizing data cleaning efforts only when necessary, thereby streamlining the analysis process. We can use daily returns directly without data cleaning steps as it is already the percent difference of the market prices. After data cleaning, all other variables are stationary and usable in the Granger causality analysis.

4.3 Granger Causality Analysis

Throughout this study, the term "causality" is employed in the context of the Granger causality test, which assesses temporal precedence and predictive relationships between variables. It does not imply causal inference or directionality in the traditional sense but rather signifies statistical associations identified through the Granger causality framework.

We conducted Granger causality tests for all pairs of features to identify significant causal relationships. For each pair of features, we obtained the lowest p-value from the Granger causality tests which the null hypothesis stands for 'Feature 2' does not Granger causes 'Feature 1' performed on a lag range between 0 and 50 days. Figure 6 shows p-values were then visualized using a heatmap representation, where each cell in the matrix corresponds to a pair of features, darker shades of blue in the matrix indicate lower p-values, suggesting stronger evidence of causality between the corresponding features. The number figuring in each cell is the lag where the best p-value has been obtained.

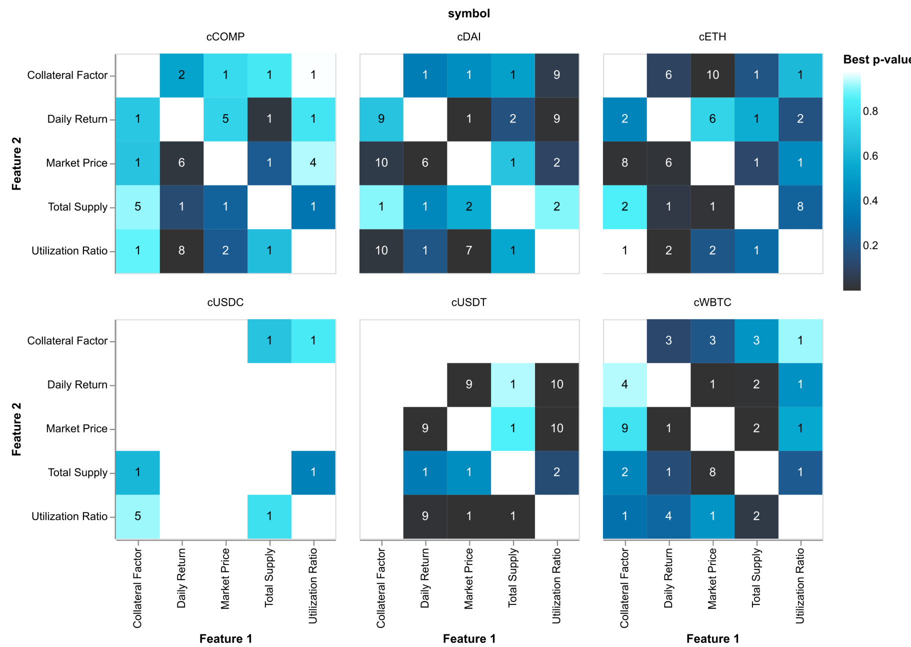


Figure 6: Granger causality test with the best p-values cross features

The heatmaps reveal varying degrees of causal relationships among the cryptocurrency market features. For instance, we observed significant "causal" links: Market Price causing Daily Return and Utilization Ratio causing Daily Return.

The variations in the patterns of causality matrix across different coins suggest that the causal relationships between pairs of features can differ substantially between cryptocurrencies. Such heterogeneity underscores the need for nuanced and coin-specific analysis, for instance, it appears that Total Supply best explains the Utilization Ratio for the volatile coins (COMP, ETH, BTC) whereas it is not the case for stablecoins (DAI and USDT).

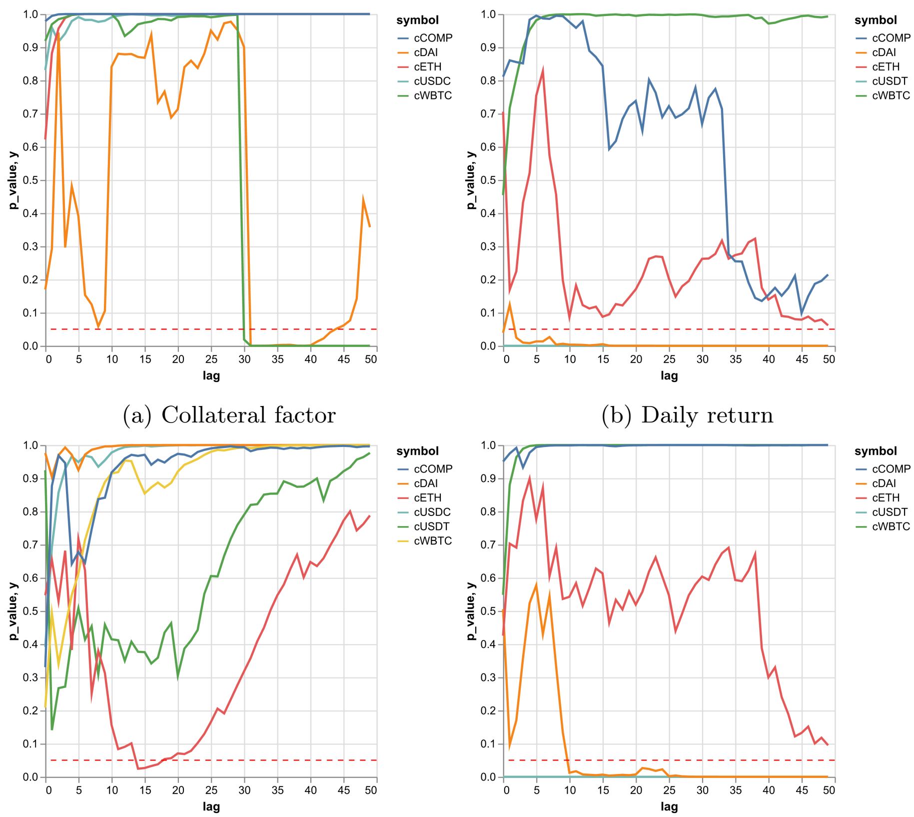
The asymmetry along the diagonal of the matrix indicates that one feature predominantly explains or influences another rather than a bidirectional relationship. In many instances, we found that the utilization ratio causes the market price but not vice versa. This directional causality provides information on the directional flow of information or influence within the cryptocurrency ecosystem.

We also identified cases where the same p-value is observed from one side to another of the diagonal, indicating mutual Granger causality between pairs of features. This suggests a bidirectional relationship, where changes in one feature influence changes in another, and reciprocally.

Next, we delved deeper into the Granger causality relationships between the previous features studied and the utilization ratio by analyzing the evolution of p-values for the Granger causality test across various time lags, ranging from

0 to 50 days. The limited number of coins that exhibit p-values below the conventional threshold of 0.05 indicates significant Granger causality between features and the utilization ratio.

Figure 7 shows that only a few coins (ETH and DAI) demonstrate p-values below 0.05 at certain time lags. Among these coins, Ethereum (ETH) emerges as the most prominent, with p-values consistently dropping near or below the 0.05 threshold for total supply, market price , and daily return. However, the majority of coins in our study exhibit p-values above the 0.05 threshold, indicating weak or insignificant Granger causality between features and the utilization ratio. This observation highlights the complexity and multifactorial nature of utilization behavior in the cryptocurrency market, wherein various factors may contribute to utilization ratio patterns beyond the features considered in our analysis.



(a) Collateral factor

(b) Daily return

(c) Total supply

(d) Market price

Figure 7: Granger causality test of features causing utilization ratio on the lag of 0 to 50 days

Further exploration through event analysis, utilizing specialized API data such as Coindar as demonstrated in the appendix 13, could unveil new insights.

5 Borrowing Activities

Taking inspiration from Saengchote's work [5] about user behavior, we decided to model the net borrowing activity on the compound instead of the total Borrowing. Since most of the borrowing activity on the platform is regrouped in the DAI, WBTC, ETH, and USDC (stablecoin whose value is pegged to the dollar) pools (see Figure 8 below), we chose to focus on these four pools.

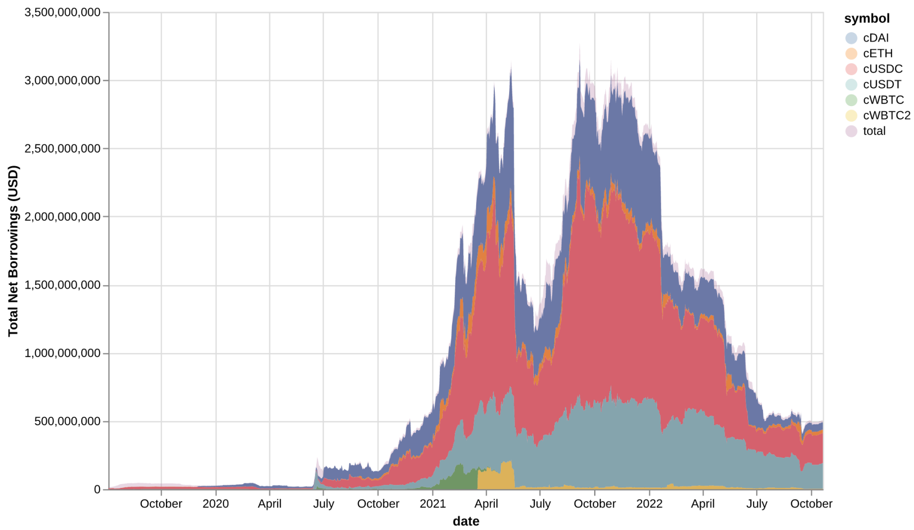


Figure 8: Total net borrowing amount in US Dollars

5.1 Model Implementation

We modeled the factors that could affect the borrowing activities. Let $\%Y_t = \frac{Y_t - Y_{t-1}}{Y_t}$ which is the percentage change of variable Y at time t. Our variable of interest is $\%totalBorrows_t$, the percentage of total borrow changes. In this section, we study the net borrowing on each pool, meaning we remove the self-borrowing amount from the actual borrowing, using different collateral assets.

Due to the nature of the data, assuming independence between the observations is unreasonable, we expect the price of ETH and the activity of day t to be correlated to day $t-1$ values. Because of that, a time series model is used. The **Autoregressive Distributed Lag (ARDL) model** is a direct extension of the VAR model that allows simultaneity: Y_t can depend on a series X_t (and not only the lagged values).

In the ARDL model, explanatory variables are selected by evaluating their correlation with the response variable. For example, the borrowing rate ($borrowRate$)

of ETH is correlated with the borrowing activity in the ETH pool. In this case, the selected variables are $\%Price(ETH)$, $borrowRate$ and $volatility7D$, $GasFee$ (market price of ETH/BTC, the borrowing rate of the pool and 7-day volatility of the ETH/BTC price, and gas fee, respectively). We use the Box-Jenkins method to test that time series are differentiated and stationary, as most of the variables are already converted into the percent difference.

We select the number of days lags: if Y_t is highly correlated with X_t, X_{t-2} but not X_{t-1}, we still use all the coefficients X_t, X_{t-1}, X_{t-2}. For each explanatory variable, we then select the maximum lags using correlation coefficients. Next, we choose which combination of lags to use in the model. A consequent number of models are fitted; the selected model is the one with the best (lowest) AIC. For example, we fit the model with lag 0 of the borrowing rate, lag 0, and 1 of price and fit another model with lag 0, 1 of rate, and lag 0 of price; we then compare the AIC.

The borrowing rate is determined by a function of the utilization ratio (i.e., the ratio between the total borrows and total lending). In the ARDL models, the borrowing rate at time t is deliberately omitted as it is a direct function of the total borrows. The equation for WBTC is slightly different; we use the price and volatility of BTC as explanatory variables instead of the price of ETH.

5.2 ARDL Regression Results

The results of the regressions are presented in Table 2 below. If the coefficient is left blank, the variable has not been selected for the corresponding pool:

Table 2: Estimating total borrows on day t

Variable	**ETH**	**USDC**	**DAI**	**WBTC**
$\%totalBorrows_{t-1}$	-0.056**	0.08***		
	(0.027)	(0.025)		
$\%totalBorrows_{t-2}$		0.098***		
		(0.024)		
$\%Price_{t}$	-0.391***	0.224***	0.160***	-0.036
	(0.146)	(0.036)	(0.038)	(0.177)
$\%Price_{t-1}$		0.114**	0.113***	0.30*
		(0.031)	(0.038)	(0.178)
$\%Price_{t-2}$		0.080***		-0.326*
		(0.031)		(0.177)
$borrowRate_{t-1}$	1.637	-0.08*	-0.337*	0.344
	(5.476)	(0.04)	(0.188)	(0.863)
$borrowRate_{t-2}$	-5.640		0.202	
	(15.443)		(0.189)	
$volatility7D$	0.276	-9.42×10^{-5} ***	-7.27×10^{-5} **	-3.21×10^{-6}
	(0.285)	(2.5×10^{-5})	(2.96×10^{-5})	(7.57×10^{-6})
$GasFee$	32.422	5.953	6.452	
	(39.746)	(9.994)	(12.248)	
$constant$	0.103**	0.011***	0.012***	-0.002
	(0.05)	(0.003)	(0.004)	(0.029)
AIC	279.680	-4874.21	-4864.996	-543.43
N	1638	1635	1627	564
MSE	0.069	0.0029	0.0046	0.02

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$, standard errors in parentheses

ETH pool: The $Price_t$ is significant in the model; this gives us proof that the price of ETH significantly impacts the outcome of the ETH pool. Volatility, borrow rate, and gas fees are although not at all significant. From the results of this regression, we can conclude that a change of 1% in the price of ETH is correlated to a drop of -0.391% in the borrowing of ETH. Intuitively, it makes sense because it is pricier to borrow ETH all other things equal. The same reasoning stands for the coefficient $borrowRate_{t-1}$: if the borrowing rate was higher the day before, people would borrow less ETH today.

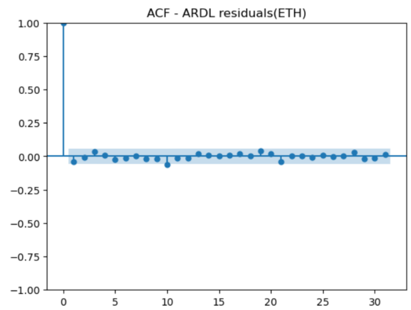
USDC pool: Tovanich et al. reported that the majority of borrowers used ETH as collateral to borrow USDC [7], that's why the price of ETH is used to explain the activity on the USD pool. All the chosen variables have low p-values, except for gas fees. The model shows that a higher price at date t (1% increase) leads to an increase in the borrowing of USDC (0.19%). The intuition is that it is more profitable for users to use ETH as collateral to borrow USDC and all in all, they can borrow more USDC with less collateral. The coefficients are also positive for $t-1$ and $t-2$. Moreover, the borrowing rate at date $t-1$ is negatively (and significantly) correlated with borrowing activity, which is not surprising since a higher borrowing rate will deter people from borrowing the asset. Although the coefficient of the volatility is low, it is still significant. An increasing volatility of the price of ETH is correlated with a slight decrease in the total borrows of USDC. A possible explanation is that the more the price of ETH is volatile, the less reliable it is as collateral. In this case, users are reticent to borrow more USDC using ETH as a collateral asset.

DAI pool: Gas fee and borrow rate (from 2 days ago) are not significant in the model, while the price of ETH (t, $t-1$, $t-2$) and borrowing rate at date $t-1$ are significant. We can interpret that a higher price at date t (1% increase) leads to an increase in the borrowing of DAI (0.16%). Borrowing DAI using ETH represents a significant part of the activity in the compound platform. Similar to the result from the USDC pool, we can assume that since ETH is pricier, it is more valuable as a collateral for borrowing DAI. The borrowing rate ($t-1$) is negative with a significant at 10% level. As expected, a high borrowing rate is correlated with a decrease in activity.

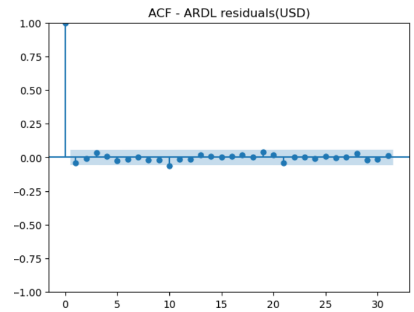
WBTC pool: The estimates of the coefficients for all variables have a high p-value, even those who are significant are significant at the 10% level (market price at date ($t-1$) and ($t-2$)). Since the modeling and regression methodology is the same for all pools; this issue may be due to a model misspecification and unobserved bias (i.e., explanatory variables that we don't have in our dataset or that we can't measure). As a result, the regression result for WBTC is inconclusive.

Figures 9 show the autocorrelation of the residual for each regression model. The three validated models passed the Ljung-Box test for residual autocorrelation, thus validating the independence of the residuals and strengthening our confidence in the modeling.

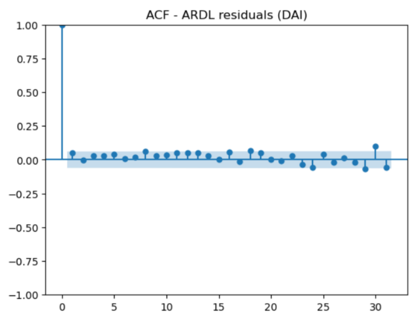
The ARDL models succeeded in identifying a significant effect of the price, and the sign of the coefficients makes sense intuitively, as explained above. Overall, we were able to find patterns in the borrowing activity. Nonetheless, the model for WBTC is the only one that is inconclusive because the coefficients are not significant.



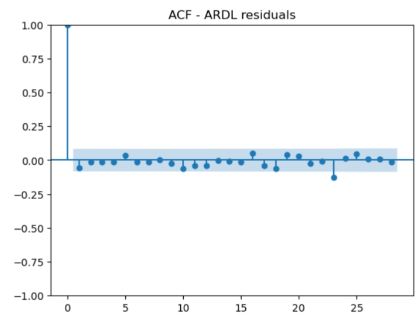
(a) ETH



(b) USD



(c) DAI



(d) WBTC

Figure 9: ACF for the ETH, USD, DAI, WBTC ARDL residuals

6 Liquidity Mining Activity

As of May 11, 2024, Compound facilitated borrowing totaling over $868,412,798, supported by $2,196,702,219 in collateral [5]. Although these figures may appear substantial, a significant portion stems from liquidity mining, particularly prominent in the early stages of the protocol, as highlighted in the research by Park and Stinner [4]. Addressing this phenomenon is crucial for maintaining the protocol's health, necessitating adjustments to liquidity programs to mitigate liquidity mining. Our study seeks to gain deeper insight into the motivations behind this user behavior. We assume that liquidity mining is driven by the following features:

- **Protocol Popularity:** We hypothesize that the popularity of the protocol is a significant factor driving liquidity mining activities. Higher participation from DeFi enthusiasts should correlate with increased liquidity mining occurrences.
- **COMP Token Price:** We assume the impact of the COMP token price on liquidity mining incentives. A higher token price translates to more

[5] https://compound.finance/

valuable rewards, further incentivizing participation.

- **Daily Distribution of COMP Tokens:** We explore how the daily distribution of COMP tokens influences liquidity mining. A greater distribution may attract more yield-seekers.

To contextualize our investigation, we initially sought to replicate the findings of Park and Stinner using our data set. Figure 10 shows the different time series used in the liquidity mining analysis.

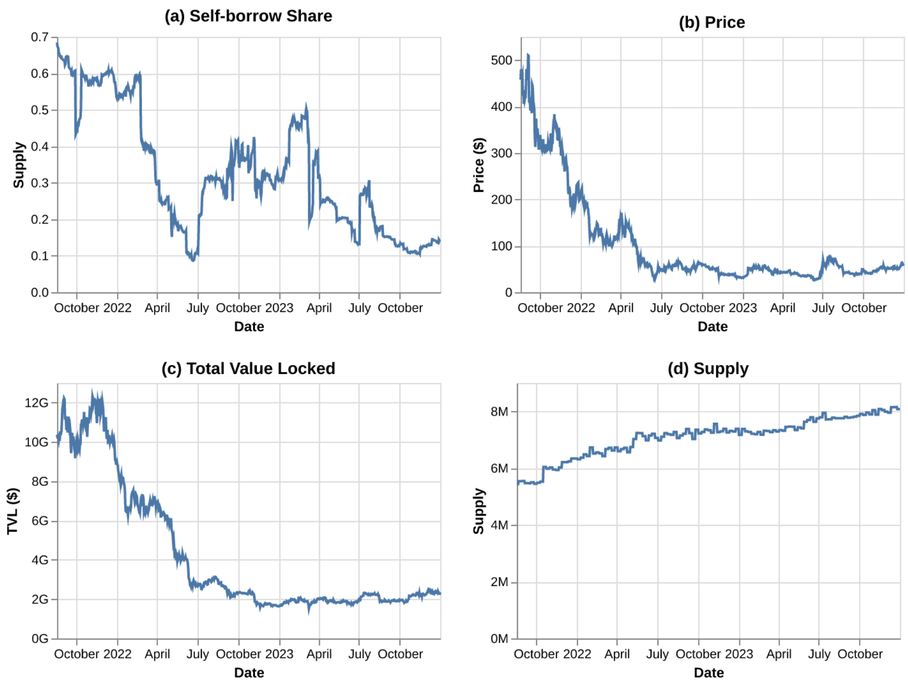


Figure 10: Time series visualization

Our analysis corroborates the findings of Park and Stinner. Figure 10 (a) shows that in October 2021, liquidity mining activities accounted for up to 85% of Compound transactions. This underscores the prevalence of liquidity mining within the protocol and highlights the need for further investigation into its impact on platform dynamics.

6.1 Time Series Assumptions for VECM model

Before embarking on our analysis, let's define the variables we'll be examining:

- (Y_t): The share of self-borrow transactions within Compound.
- $(X1_t)$: The price of COMP token in dollars.

- ($X2_t$): The total value locked (TVL) of Compound.
- ($X3_t$): The total supply of COMP tokens.

To ensure the suitability of our analysis, we must verify certain assumptions, particularly regarding the stationarity and cointegration of our variables. Firstly, stationarity is a critical assumption for time-series analysis. To assess stationarity, we examine the first differences of our variables, denoted as ΔY, $\Delta X1$, $\Delta X2$, and $\Delta X3$. The Augmented Dickey-Fuller (ADF) test, presented in Table 3, indicates that all the first differences are stationary at a significance level of 1%.

Table 3: ADF test results

	ΔY	$\Delta X1$	$\Delta X2$	$\Delta X3$
ADF Statistic 1%: -3.438 5%: -2.865 10%: -2.569	-28.9	-6.1	-5.1	-15.3
p-value	0.0	8.5e-08	1.6e-05	4.5e-28

Next, we assess cointegration among our variables using the Johansen Cointegration Test. The results, presented in Table 4, indicate co-integration among variables, as evidenced by trace statistics exceeding the critical values at the significance levels 1%, 5%, and 10%.

Table 4: Johansen Cointegration Test

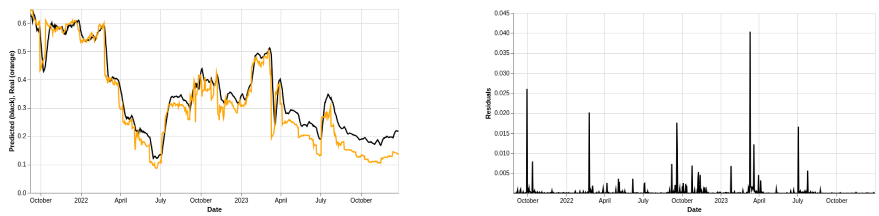
	ΔY	$\Delta X1$	$\Delta X2$	$\Delta X3$
Trace Statistic	1619.0	1020.4	660.0	301.3
1%	54.7	35.5	20.0	6.6
5%	47.9	29.8	15.5	3.8
10%	44.5	27.1	13.4	2.7

Based on these results, we can proceed with confidence in using the **Vector Error Correction Model (VECM)** for our analysis, as it is suitable for modeling the dynamic relationships among cointegrated variables in a time series framework.

6.2 Vector Error Correction Model (VECM) Results

Having fitted the VECM model to our dataset, we obtained insightful results regarding the dynamics of liquidity mining within Compound. Table 5 in the appendix summarizes the model coefficients for the equation ΔY_t, which represents the change in the share of self-borrow transactions within Compound.

Additionally, we visualized the model's performance through two key visualizations, presented below.



(a) True vs. Predicted ΔY_t

(b) Squared Residuals

Figure 11: Model Performance Metrics

In Figure 11a, we visualize the true values of ΔY_t (depicted in orange) and the predicted values generated by the VECM model (depicted in black) over time. It seems like the model captured the trend of the self-borrow share over time but a further investigation has to be made to confirm quantitatively this alignment.

Figure 11b illustrates the squared residuals, calculated as the squared difference between the true and predicted values of ΔY_t. The proximity of the majority of residuals to zero suggests that the model performs well in most instances. However, occasional spikes in the residuals indicate discrepancies between the true and predicted values, which may be attributed to outliers present in the dataset. However, we compared the mean squared error and it is almost equal to the variance of ΔY_t so our model does not improve upon using the mean of ΔY_t as a predictor. Despite this, the model can still provide insights into the relationships between the exogenous variables and the dependent variable.

To ensure the reliability of our model, we further analyzed the autocorrelation function of the residuals. Figure 12 presents the autocorrelation function for different lags, revealing a significant spike at lag 7, but it is not quite enough for the Ljung-Box test to be significant at the 5% level therefore indicating no autocorrelation between variables. In any case, the autocorrelation is not particularly large, and at lag 7 it is unlikely to have any noticeable impact on the forecasts or the prediction intervals.

Looking at Table 2, we can interpret the coefficients of the ΔY_t equation as follows:

- The coefficients of ΔY_{t-i} for $i \in 1, \ldots, 7$ are significant at the 5% level. This indicates that past values of the self-borrow share, ΔY_t, have a statistically significant impact on its current value. Notably, all these coefficients are negative, suggesting a negative autocorrelation. This means that an increase in the self-borrow share in the previous week tends to be followed by a decrease in the current week's share, implying a potential

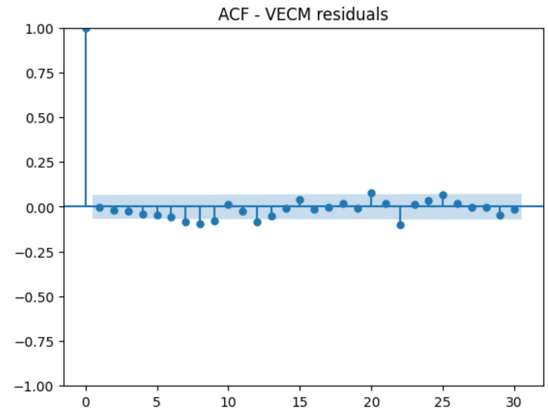


Figure 12: Residuals Autocorrelation Function For Different Lags

corrective mechanism or mean-reverting behavior in the self-borrow share over time.

- Both coefficient of $\Delta X1_{t-6}$ and $\Delta X3_{t-2}$ are significant but really close to zero. Their small magnitude suggests that the immediate past value of $\Delta X1_t$ and $\Delta X3_t$ has a negligible effect on the current value of ΔY_t. This could indicate that although there is a detectable relationship, the economic or practical significance of this effect is minimal.

- The alpha coefficient associated with the equation is not significant at the 5% level so we can assume that the variables can deviate from the long-run equilibrium depicted in table 9: $\Delta Y_t - 0.0042\Delta X1_t - 1.517e^{-06}\Delta X3_t$ is stationary.

Overall, the VECM model shows that the exogenous variables have a limited impact on the self-borrow share over time. However, the data shows a potential corrective mechanism in the self-borrow share over time.

7 Conclusion

Our statistical analysis underscores the nuanced nature of causal relationships between users' behaviors in the Compound lending protocol and the cryptocurrency market. The Granger causality results reveal that the causal relationships between market features vary substantially between different cryptocurrencies. Notably, total supply best explains the utilization ratio for volatile coins like COMP, ETH, and BTC, whereas this is not the case for stablecoins such as DAI and USDT. Ethereum (ETH) emerges as a significant player, demonstrating consistent and significant causal links with total supply, market price, and daily return.

The asymmetry observed along the diagonal of the causality matrix suggests directional causality within the cryptocurrency ecosystem. For instance, the utilization ratio causes the market price in many instances, indicating a unidirectional relationship. Additionally, cases of mutual Granger causality between feature pairs indicate bidirectional relationships, where changes in one feature influence changes in another, and reciprocally. The distinction between volatile and stablecoins is evident in the observed causal relationships. Volatile coins like COMP, ETH, and BTC exhibit dynamic causal relationships due to their price volatility, while stablecoins like DAI and USDT show more predictable causal dynamics.

To thoroughly study the relationship between the price, borrow rate, volatility, and net borrowing activity, we chose to fit a time series regression model. The four major pools are studied and the variable interest is the change in total net borrowings. We found that if the price of the collateral (ETH) is higher, people will borrow more DAI and USDC in repercussion. The volatility of the price is, on the contrary, negatively correlated with the borrowing activities. The opposite happens when trying to borrow ETH: if the price increases, it results in a drop in borrowing activities.

Finally, we also assess the dynamics of liquidity mining within the Compound lending protocol. We began by revisiting Park and Stinner's findings, which highlight the significant role of liquidity mining in the Compound's ecosystem. Under the correct assumptions, the Vector Error Correction Model (VECM) revealed that the impact of Compound popularity, the COMP price, and the token supply is limited on the self-borrow share. Over time, the model has highlighted a potential corrective mechanism or mean-reverting behavior. While exogenous variables such as the price of the COMP token and the total supply of COMP tokens have a limited impact on the self-borrow share, our findings underscore the prevalence of liquidity mining within the protocol and highlight the need for further investigation into the usages of DeFi platforms.

8 Bibliography

[1] Jean Barthélemy, Clément Delaneau, Thomas Martignon, Benoît Nguyen, Marten Riho Pallum, and Salim Talout Zitan. Interest rates in decentralised finance. Technical report, 2024.

[2] Robert Leshner and Geoffrey Hayes. Compound: The money market protocol. Technical report, 2019.

[3] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.

[4] Andreas Park and Jona Stinner. Phantom liquidity in decentralized lending. *Available at SSRN 4489732*, 2023.

[5] Kanis Saengchote. Decentralized lending and its users: Insights from compound. *Journal of International Financial Markets, Institutions and Money*, 87:101807, 2023.

[6] Fabian Schär. Decentralized finance: On blockchain-and smart contract-based financial markets. *FRB of St. Louis Review*, 2021.

[7] Natkamon Tovanich, Myriam Kassoul, Simon Weidenholzer, and Julien Prat. Contagion in decentralized lending protocols: A case study of compound. In *Proceedings of the 2023 Workshop on Decentralized Finance and Security*, pages 55–63, 2023.

A Liquidity Mining VECM results

Table 5: Deterministic terms outside the cointegration relation parameters for equation selfBorrowShare (ΔY_t)

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
L1.selfBorrowShare	-0.8250	0.040	-20.422	0.000	-0.904	-0.746
L1.priceUSD	1.136e-06	0.000	0.008	0.993	-0.000	0.000
L1.TVL	2.304e-12	3.91e-12	0.589	0.556	-5.36e-12	9.97e-12
L1.supply	-6.157e-08	3.38e-08	-1.821	0.069	-1.28e-07	4.7e-09
L2.selfBorrowShare	-0.7103	0.048	-14.806	0.000	-0.804	-0.616
L2.priceUSD	5.009e-06	0.000	0.031	0.975	-0.000	0.000
L2.TVL	-3.472e-12	6.22e-12	-0.558	0.577	-1.57e-11	8.72e-12
L2.supply	-7.064e-08	3.14e-08	-2.247	0.025	-1.32e-07	-9.02e-09
L3.selfBorrowShare	-0.5754	0.052	-11.082	0.000	-0.677	-0.474
L3.priceUSD	0.0001	0.000	0.707	0.480	-0.000	0.000
L3.TVL	-4.029e-14	7.71e-12	-0.005	0.996	-1.51e-11	1.51e-11
L3.supply	-5.552e-08	2.89e-08	-1.919	0.055	-1.12e-07	1.18e-09
L4.selfBorrowShare	-0.4816	0.053	-9.169	0.000	-0.585	-0.379
L4.priceUSD	0.0001	0.000	0.835	0.404	-0.000	0.000
L4.TVL	-2.92e-12	8.22e-12	-0.355	0.722	-1.9e-11	1.32e-11
L4.supply	-3.99e-08	2.63e-08	-1.520	0.129	-9.13e-08	1.16e-08
L5.selfBorrowShare	-0.3411	0.051	-6.725	0.000	-0.441	-0.242
L5.priceUSD	0.0002	0.000	1.160	0.246	-0.000	0.000
L5.TVL	-3.342e-12	8.09e-12	-0.413	0.679	-1.92e-11	1.25e-11
L5.supply	-2.585e-08	2.31e-08	-1.119	0.263	-7.11e-08	1.94e-08
L6.selfBorrowShare	-0.2114	0.045	-4.677	0.000	-0.300	-0.123
L6.priceUSD	0.0003	0.000	2.058	0.040	1.22e-05	0.000
L6.TVL	-6.876e-12	7.05e-12	-0.975	0.330	-2.07e-11	6.95e-12
L6.supply	-1.676e-08	1.93e-08	-0.869	0.385	-5.45e-08	2.1e-08
L7.selfBorrowShare	-0.1086	0.034	-3.159	0.002	-0.176	-0.041
L7.priceUSD	0.0001	7.63e-05	1.635	0.102	-2.48e-05	0.000
L7.TVL	-8.002e-12	5.1e-12	-1.568	0.117	-1.8e-11	2e-12
L7.supply	2.725e-09	1.41e-08	0.193	0.847	-2.5e-08	3.05e-08
ec1	-0.0314	0.025	-1.259	0.208	-0.080	0.017

Table 6: Deterministic terms outside the cointegration relation parameters for equation priceUSD ($\Delta X1_t$)

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
L1.selfBorrowShare	-17.1548	13.393	-1.281	0.200	-43.405	9.096
L1.priceUSD	-0.8756	0.046	-19.169	0.000	-0.965	-0.786
L1.TVL	3.276e-08	1.3e-09	25.258	0.000	3.02e-08	3.53e-08
L1.supply	2.351e-05	1.12e-05	2.097	0.036	1.54e-06	4.55e-05
L2.selfBorrowShare	-29.5553	15.905	-1.858	0.063	-60.729	1.618
L2.priceUSD	-0.8320	0.053	-15.619	0.000	-0.936	-0.728
L2.TVL	3.395e-08	2.06e-09	16.466	0.000	2.99e-08	3.8e-08
L2.supply	2.134e-05	1.04e-05	2.047	0.041	9.11e-07	4.18e-05
L3.selfBorrowShare	-39.1816	17.215	-2.276	0.023	-72.922	-5.441
L3.priceUSD	-0.7279	0.058	-12.605	0.000	-0.841	-0.615
L3.TVL	3.123e-08	2.56e-09	12.222	0.000	2.62e-08	3.62e-08
L3.supply	2.271e-05	9.59e-06	2.368	0.018	3.91e-06	4.15e-05
L4.selfBorrowShare	-20.2426	17.413	-1.163	0.245	-54.371	13.885
L4.priceUSD	-0.4971	0.057	-8.658	0.000	-0.610	-0.385
L4.TVL	2.966e-08	2.72e-09	10.886	0.000	2.43e-08	3.5e-08
L4.supply	2.114e-05	8.7e-06	2.429	0.015	4.08e-06	3.82e-05
L5.selfBorrowShare	-13.9274	16.818	-0.828	0.408	-46.889	19.034
L5.priceUSD	-0.2416	0.052	-4.634	0.000	-0.344	-0.139
L5.TVL	1.987e-08	2.68e-09	7.411	0.000	1.46e-08	2.51e-08
L5.supply	1.022e-05	7.66e-06	1.335	0.182	-4.79e-06	2.52e-05
L6.selfBorrowShare	1.7740	14.986	0.118	0.906	-27.598	31.146
L6.priceUSD	-0.1659	0.041	-4.041	0.000	-0.246	-0.085
L6.TVL	9.66e-09	2.34e-09	4.131	0.000	5.08e-09	1.42e-08
L6.supply	7.913e-06	6.39e-06	1.238	0.216	-4.61e-06	2.04e-05
L7.selfBorrowShare	6.1204	11.399	0.537	0.591	-16.221	28.462
L7.priceUSD	-0.1273	0.025	-5.028	0.000	-0.177	-0.078
L7.TVL	5.723e-09	1.69e-09	3.382	0.001	2.41e-09	9.04e-09
L7.supply	2.338e-06	4.69e-06	0.499	0.618	-6.85e-06	1.15e-05
ec1	17.5863	8.260	2.129	0.033	1.396	33.776

Table 7: Deterministic terms outside the cointegration relation parameters for equation TVL ($\Delta X2_t$)

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
L1.selfBorrowShare	-6.167e+08	3.5e+08	-1.764	0.078	-1.3e+09	6.86e+07
L1.priceUSD	1.717e+06	1.19e+06	1.440	0.150	-6.2e+05	4.05e+06
L1.TVL	-0.8614	0.034	-25.441	0.000	-0.928	-0.795
L1.supply	1018.1687	292.643	3.479	0.001	444.599	1591.738
L2.selfBorrowShare	-4.728e+08	4.15e+08	-1.139	0.255	-1.29e+09	3.41e+08
L2.priceUSD	2.283e+06	1.39e+06	1.642	0.101	-4.43e+05	5.01e+06
L2.TVL	-0.8334	0.054	-15.482	0.000	-0.939	-0.728
L2.supply	735.0821	272.078	2.702	0.007	201.820	1268.345
L3.selfBorrowShare	-4.171e+08	4.49e+08	-0.928	0.353	-1.3e+09	4.64e+08
L3.priceUSD	2.506e+06	1.51e+06	1.663	0.096	-4.48e+05	5.46e+06
L3.TVL	-0.6631	0.067	-9.938	0.000	-0.794	-0.532
L3.supply	643.4246	250.357	2.570	0.010	152.735	1134.115
L4.selfBorrowShare	-7.401e+08	4.55e+08	-1.628	0.104	-1.63e+09	1.51e+08
L4.priceUSD	5.111e+04	1.5e+06	0.034	0.973	-2.89e+06	2.99e+06
L4.TVL	-0.5271	0.071	-7.411	0.000	-0.667	-0.388
L4.supply	530.9041	227.190	2.337	0.019	85.620	976.188
L5.selfBorrowShare	-9.346e+08	4.39e+08	-2.129	0.033	-1.8e+09	-7.41e+07
L5.priceUSD	-7.344e+05	1.36e+06	-0.539	0.590	-3.4e+06	1.93e+06
L5.TVL	-0.3824	0.070	-5.464	0.000	-0.520	-0.245
L5.supply	517.3347	199.913	2.588	0.010	125.513	909.157
L6.selfBorrowShare	-5.467e+08	3.91e+08	-1.398	0.162	-1.31e+09	2.2e+08
L6.priceUSD	-3.798e+06	1.07e+06	-3.545	0.000	-5.9e+06	-1.7e+06
L6.TVL	-0.2089	0.061	-3.422	0.001	-0.329	-0.089
L6.supply	539.0861	166.844	3.231	0.001	212.078	866.094
L7.selfBorrowShare	-2.936e+07	2.98e+08	-0.099	0.921	-6.13e+08	5.54e+08
L7.priceUSD	-2.193e+06	6.61e+05	-3.318	0.001	-3.49e+06	-8.98e+05
L7.TVL	-0.0370	0.044	-0.837	0.403	-0.124	0.050
L7.supply	378.3897	122.445	3.090	0.002	138.403	618.377
ec1	7.58e+08	2.16e+08	3.515	0.000	3.35e+08	1.18e+09

Table 8: Deterministic terms outside the cointegration relation parameters for equation supply ($\Delta X3_t$)

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
L1.selfBorrowShare	-6.676e+05	9.51e+04	-7.020	0.000	-8.54e+05	-4.81e+05
L1.priceUSD	2862.3458	324.362	8.825	0.000	2226.608	3498.083
L1.TVL	-1.42e-05	9.21e-06	-1.542	0.123	-3.23e-05	3.85e-06
L1.supply	0.2012	0.080	2.528	0.011	0.045	0.357
L2.selfBorrowShare	-5.864e+05	1.13e+05	-5.192	0.000	-8.08e+05	-3.65e+05
L2.priceUSD	2335.7453	378.276	6.175	0.000	1594.338	3077.153
L2.TVL	-9.828e-06	1.46e-05	-0.671	0.502	-3.85e-05	1.89e-05
L2.supply	0.2248	0.074	3.038	0.002	0.080	0.370
L3.selfBorrowShare	-4.664e+05	1.22e+05	-3.815	0.000	-7.06e+05	-2.27e+05
L3.priceUSD	2450.4093	410.097	5.975	0.000	1646.635	3254.184
L3.TVL	-6.747e-06	1.81e-05	-0.372	0.710	-4.23e-05	2.88e-05
L3.supply	0.2246	0.068	3.298	0.001	0.091	0.358
L4.selfBorrowShare	-4.827e+05	1.24e+05	-3.904	0.000	-7.25e+05	-2.4e+05
L4.priceUSD	2445.4289	407.701	5.998	0.000	1646.350	3244.508
L4.TVL	-2.986e-05	1.93e-05	-1.543	0.123	-6.78e-05	8.06e-06
L4.supply	0.2349	0.062	3.801	0.000	0.114	0.356
L5.selfBorrowShare	-3.261e+05	1.19e+05	-2.731	0.006	-5.6e+05	-9.21e+04
L5.priceUSD	1448.9032	370.316	3.913	0.000	723.096	2174.710
L5.TVL	-5.012e-05	1.9e-05	-2.632	0.008	-8.74e-05	-1.28e-05
L5.supply	0.2550	0.054	4.688	0.000	0.148	0.362
L6.selfBorrowShare	-5.951e+04	1.06e+05	-0.559	0.576	-2.68e+05	1.49e+05
L6.priceUSD	1123.2058	291.508	3.853	0.000	551.861	1694.551
L6.TVL	-8.691e-06	1.66e-05	-0.523	0.601	-4.12e-05	2.39e-05
L6.supply	0.2601	0.045	5.732	0.000	0.171	0.349
L7.selfBorrowShare	-1459.0283	8.09e+04	-0.018	0.986	-1.6e+05	1.57e+05
L7.priceUSD	603.7413	179.731	3.359	0.001	251.474	956.008
L7.TVL	3.374e-06	1.2e-05	0.281	0.779	-2.02e-05	2.69e-05
L7.supply	-0.0300	0.033	-0.902	0.367	-0.095	0.035
ec1	7.813e+05	5.87e+04	13.320	0.000	6.66e+05	8.96e+05

Table 9: Long run relations between variables.

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
beta.1	1.0000	0	0	0.000	1.000	1.000
beta.2	-0.0042	0.001	-4.042	0.000	-0.006	-0.002
beta.3	-2.863e-11	4.61e-11	-0.622	0.534	-1.19e-10	6.16e-11
beta.4	-1.517e-06	1.17e-07	-13.011	0.000	-1.75e-06	-1.29e-06

B Major Events Influence

Leveraging the CoinDar API, we tried to incorporate chronological events into our analysis to assess the relevance and impact of major events (regulation changes, protocol governance updates) on coin features such as market price or utilization ratios as shown in Figure 13.

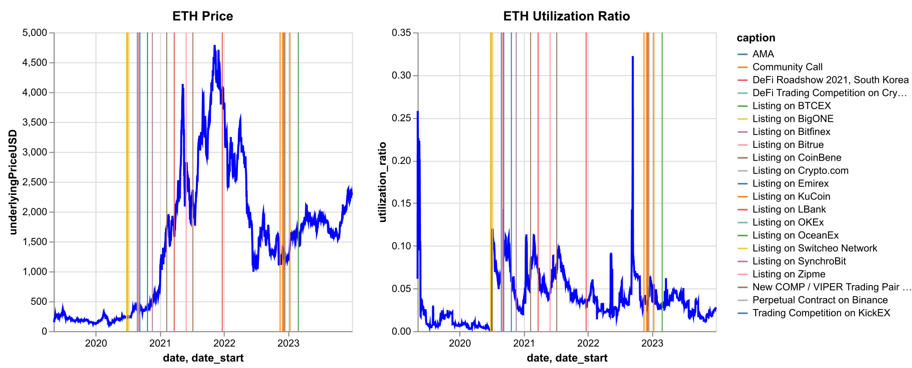


Figure 13: Evolution of BTC market price and utilization ratio in perspective of important events

However, we did not pursue any quantitative analysis but one might use it as a starting lead to dig further into Compound's user behavior.