Applied Statistics Project

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

Ulysse Kazmierczak, Tsitohaina Ravelomanana, Bilal Benhana

ENSAE Paris

May 2024

Introduction

We build our work from Natkamon Tovanich's paper: Contagion in Decentralized Lending Protocols: A Case Study of Compound.

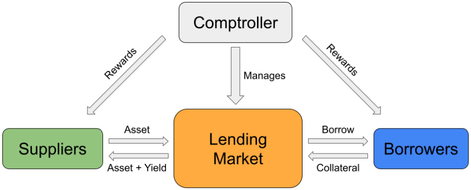


Fig. 1: Compound Visualizer (Simplified Version)

Research Questions

Our study delved deeper into three mechanisms of Compound:

- Interplay of Market and Pool Features
- Impact of Ethereum Price Fluctuations
- Liquidity Mining Practices

Data Collection

Most of our data comes from Natkamon Tovanich's balance sheets. However, we aimed to enrich our data with data aggregators like DefiLlama, CoinMarketCap, and CoinGecko.

Our dataset can be characterized as follows:

- A small number of features.
- A lot of observations over time.

We decided to use time series to handle the analysis properly.

Market Price and Utilization Ratio

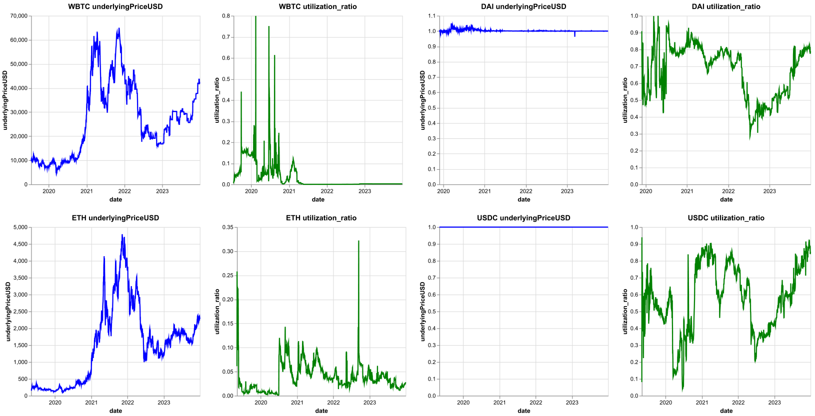


Fig. 2: Market Price vs Utilization Ratio for Stable Coin and Volatile Coin Representatives

Price and Utilization Ratio Trends

- Figure 2 illustrates the evolution of prices with their respective utilization ratios (in USD) for volatile assets (ETH, BTC) and stable coins (USDC, DAI)
- Graphically, the prices of ETH and BTC appear to be positively correlated.
- However, their utilization ratios do not exhibit a clear correlation.
- For stable coins which are pegged to $1 all the time by definition, their utilization ratio seems to show a clearer correlation.

Observations

This observation raises two possibilities:

1. Either the market price alone is not a suitable variable to explain the utilization ratio.
2. Or, a more in-depth statistical analysis is required to uncover potential relationships. (especially to uncover potential behavior differences regarding the coin category (volatile / stable) or even at the core of a same coin class)

Overview

To examine the lead/lag autocorrelation between various key features of a given coin using Granger causality testing, several crucial steps are undertaken to ensure the validity and reliability of the results.

Preparation of Time Series Data

1. **Data Normalization**
   - Normalize the time series data to standardize the range of variables. This is done by:
     - Subtracting the mean of each series from every data point.
     - Dividing by the standard deviation of the series.
2. **Trend Removal**
   - Detrend the time series to eliminate long-term trends that can bias the results. This is achieved through:
     - First differencing, which involves taking the difference between consecutive data points.
3. **Augmented Dickey-Fuller (ADF) Test**
   - Perform an ADF test to check for stationarity. The null hypothesis of this test is that the series has a unit root (i.e., it is non-stationary).

Granger Causality Testing

1. With all variables confirmed to be stationary, proceed to the Granger causality tests to investigate the potential causal relationships between:
   - Collateral factor
   - Total supply
   - Market price
   - Utilization ratio
   - Daily returns

Granger Causality Tests and Heatmap Visualization

- 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 in each cell is the lag where the best p-value has been obtained.

Granger Causality Heatmap

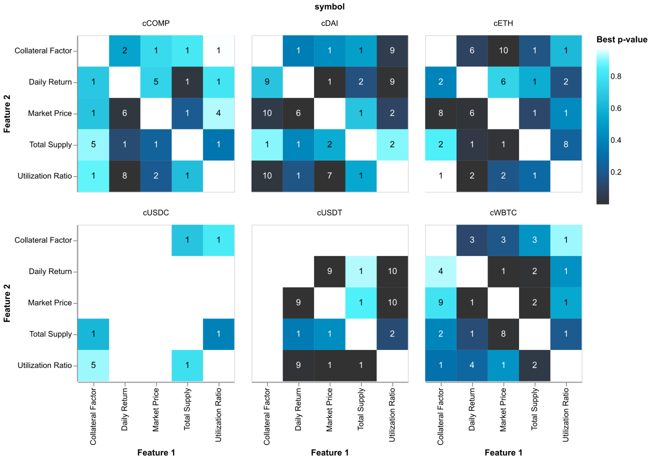


Fig. 3: Granger causality test with the best p-values cross features

Causal Relationships and Heterogeneity Across Coins

The heatmaps reveal varying degrees of causal relationships among the cryptocurrency market features.

- We observed significant 'causal' links: Market Price causing Daily Return and Utilization Ratio causing Daily Return.
- Causal relationships between pairs of features can differ substantially between cryptocurrencies.
- For instance, Total Supply best explains the Utilization Ratio for volatile coins (COMP, ETH, BTC) whereas it is not the case for stablecoins (DAI and USDT).

Directionality of Relationships

- Asymmetry along the diagonal indicates directional causality.
- Utilization Ratio often causes the daily return but not vice versa, revealing a directional flow of influence within the cryptocurrency ecosystem.
- Instantaneous granger causality (same color across the diagonal for ETH Market Price and Collateral Factor) between pairs of features indicates bidirectional relationships.(changes in one feature influence changes in another, and reciprocally)

Evolution of P-values

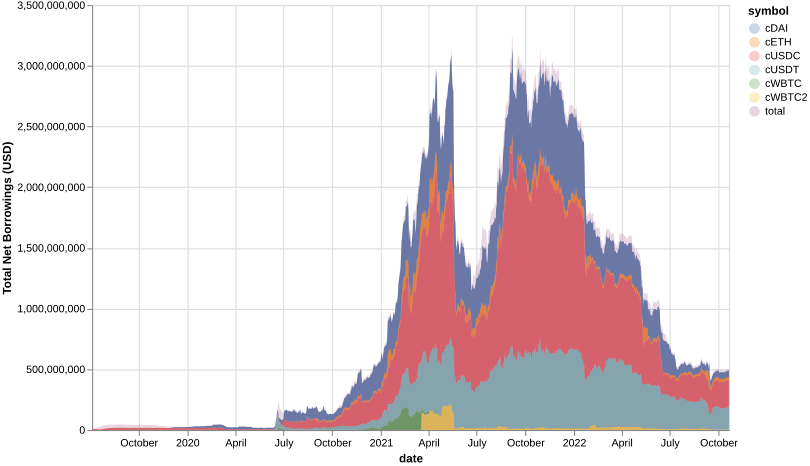
- Granger causality relationships between features and the utilization ratio were analyzed across various time lags, ranging from 0 to 50 days.
- Majority of coins exhibit p-values above the 0.05 threshold, indicating weak or insignificant Granger causality between features and the utilization ratio.
- However, Ethereum (ETH) emerges as the most prominent, with consistently dropping p-values near or below the 0.05 threshold for total supply, market price, and daily return.

Variables

- **The variable of interest** : we modeled the factors that could affect the borrowing activities. Let Y_t be a time series and let : $\%Y_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}$ which is the percentage change of variable Y at time t. Thee variable of interest is $\%totalBorrows_t$.
- **Explanatory variables** : $\%Price(ETH/USD)$, $borrowRate$ and $volatility7D$, $GasFee$, the explanatory variables are pre-selected using their correlation to the outcome

Pools

We choosed to focus on USDC, WBTC, DAI, ETH pools since most of the borrowing activity is concentrated on those.



Model

The ARDL : direct extension of the VAR model, it includes the lag 0 of the model.

$$\begin{aligned} Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \\ + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \delta_q X_{t-q} + u_t. \end{aligned}$$

Hypotheses :

- u_t is a white noise,
- the time series can be $I(1), I(0)$(integrated at order 0, 1) but not $I(2)$,
- $\mathbf{E}(u_t | Y_{t-1}, Y_{t-2}, \ldots, X_{t-1}, X_{t-2})$.

Model selection

We choose a maximum lag for each variable and then try out all possible combinations and compare the AICs. The equation for WBTC is slightly different, as price and volatility of BTC is used.

USDC: $\%totalBorrows_t = constant + \beta_1\%totalBorrows_{t-1} + \beta_2\%totalBorrows_{t-2} + \delta_0\%Price_t + \delta_1\%Price_{t-1} + \delta_2\%Price_{t-2} + \rho_1 borrowRate_{t-1} + \nu_0 volatility7D + \alpha_0 GasFee_t + \epsilon_t$

Model results

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

Interpretation

- **ETH pool:** The $\textit{Price}_t$ is significant in the model and negatively correlated with the borrowing activity. It makes sense because it is pricier to borrow ETH all other things equal. The same reasoning stands for the coefficient $\textit{borrowRate}_{t-1}$.
- **USDC pool:** a higher price at date t leads to an increase in the borrowing of USDC. The intuition is that users can borrow more USDC with less collateral. The borrowing rate at date $t-1$ is negatively (and significantly) correlated with borrowing activity and an increasing volatility of the price of ETH is correlated with a slight decrease in the total borrows of USDC.

Interpretation

- **DAI pool:** the interpretation of the coefficient of $\%$*price* is the same as in the USDC pool. The coefficient of the borrowing rate $(t-1)$ is negative.
- **WBTC pool:** The estimates of the coefficients for all variables have a high p-value, even those who are significant are significant only at the 10% level : model misspecification or endogeneity

Assumptions

To explain liquidity mining activities (Self-borrow Share) within Compound, we used three assumptions:

- Higher participation from DeFi enthusiasts should correlate with increased liquidity mining occurrences (TVL).
- A higher token price translates to more valuable rewards, further incentivizing participation (COMP price).
- A greater distribution may attract more yield-seekers (Supply).

Predictions

Under the correct assumptions, we used a VECM model to explain the relationships between the variables.

Fig. 6: True vs. Predicted Self-borrow Share

Residuals

Some outliers in the true self-borrow share lead to a slight offset between the predictions and the true share and spikes within the squared residuals.

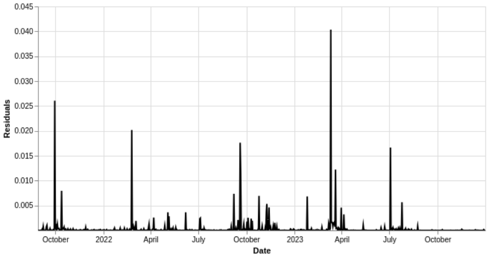


Fig. 7: Squared Residuals of the Self-borrow Share

Model Results Interpretation

The VECM model highlighted the following:

- Exogenous variables have a limited impact on the self-borrow share over time.
- The data shows a potential corrective mechanism in the self-borrow share over time.

Conclusion

Encountered difficulties:

- Get a full grasp of the Compound protocol.
- Select the correct features to explain user behavior.
- Interpret the results.
- Find the correct model.

Even though our study aims to be as exhaustive as possible, it highlights the necessity to investigate further DEFI lending protocols to get a full grasp of this new type of finance.