Analysis of the predictability power of Ethereum blockchain data

1st Giro Tomas
EPFL MFE student
Lausanne, Switzerland
tomas.girolarraz@epfl.ch

2nd Meylan Valentin
EPFL MFE student
Lausanne, Switzerland
valentin.meylan@epfl.ch

***Abstract*—We investigate if we can find causation relations between deposits to, and withdrawals from exchanges' wallets, and volatility of ERC-20 tokens (more precisely ZRX, THETA and ENJ). We find no significant evidence of predictability of volatility with linear models (OLS) and non-linear ones (Random Forest). We find significance in the predictability power of transactions from and to exchanges (for the Theta and Enjin tokens only) and bitcoin prices to predict volatility, with regressive ARIMA models.**

***Index Terms*—volatility, cryptocurrency, ERC-20, Ethereum, blockchain, prediction, machine learning, dynamic regression, ARIMAX**

INTRODUCTION

ERC tokens are digital assets that are being built on top of the Ethereum blockchain. Tether, Enjin and 0x are three examples of these tokens. Previous research by M. Griffin Shams, 2018 [1] shows that Tether issuance (i.e. creation) is linked with Bitcoin price movements. From these findings, we wondered whether on-chain transactions for ERC-20 tokens had predictive power on off-chain movements, with respect to volatility forecasting. We analyze whether inflows towards, and outflows from exchanges' wallets and bitcoin prices had predictive power on ERC-20 token volatility. We were not able to find any predictive power with Linear Regression models and with Random Forest models. On the other hand we did find significant coefficients for the volatility forecasts using regressive ARIMA models. We believe that this is due to the noisiness of the daily financial returns as explained by Black 1986 [2].

We chose to study the ERC-20 tokens ZRX, THETA and ENJ, because they one of the major ERC-20 tokens in the Ethereum blockchain. "ZRX is an Ethereum token that is used to power the 0x protocol. The protocol itself is designed to allow Ethereum tokens to be traded at a low cost directly from your wallet."[1]. "Theta's innovation is set to disrupt today's online video industry much in the same way that the YouTube platform did to traditional video back in 2005." [2] "Since its founding in 2009, Enjin has launched a gaming community platform called the Enjin Network and has grown it to approximately 20+ million users. In 2017 following an ICO that reportedly raised \$18.9 million, Enjin has been building a suite of user-centric blockchain products that aims to enable anyone to create, manage, trade, store, explore, distribute, and integrate blockchain assets." [3]

We divide our work into two sections. First, in section I. we present the data acquisition process and the engineered features. In section I.A.1 and I.A.2 we extract all tokens transactions between exchanges' wallets and other wallets. In section I.A.3 we query the price data from an API. Then in section I.B.1 we preprocess the data, and in I.B.2, we engineer the main features for our analysis such as the on-chain transaction volume and the volatility and we present the final set of variables. The section I.C presents the correlation matrices of the data.

In section II, we use a set of predicting techniques to check whether the features are statistically significant and how well they predict volatility. First, in II.A we find that all variables are statistically significant using a F-test. Second, in II.B we fit a Ridge Regression that is unable to predict the variance. Although the out of sample R^2 in our regression is low, in part II.C we compute the Linear Regression R^2 for all the features for the training set and we find that the most explanatory are the bitcoin price and the volume of transactions coming from exchanges. These variables have an R^2 of around 25% in the training data set. In section II.D we fit a Random Forest model but the results also show that it is unable to predict out of sample data. Finally in section II.E we test our features in an ARIMA dynamic regression model and we find the following results. We find evidence of predictability in the following variables: the on-chain (on the blockchain) transactions towards and from exchanges wallets, the bitcoin price and the off-chain trading volume. The results show that the relationship between an increase in variance and Bitcoin price and transactions towards exchanges is positive for some tokens. They also show that the relationship between transactions from exchanges and volatility change is negative.

[1] 0x Price Chart (ZRX), Coinbase, (1/7/2020), https://www.coinbase.com/price/0x

[2] Theta Token, https://www.thetatoken.org/

[3] Enjin Coin, https://coinmarketcap.com/currencies/enjin-coin/

I. Data Acquisition, Preprocessing and Feature Engineering

A. Data acquisition

1) On-chain transaction data: A detailed dataset aggregating the whole Ethereum blockchain is available online: all transactions for all ERC-20 tokens[4]. The dataset size is around 850 GB and contains the set of all Ethereum blocks and their attributes. The data is exported regularly using this GitHub open source code. The dataset contains a token_transfers table with the following columns:

- token_address (ERC-20 token smart contract address);
- from_address (address of the sender);
- to_address (address of the receiver);
- value (amount of tokens transferred);
- transaction_hash;
- log_index (log index in the transaction receipt);
- block_timestamp (timestamp of the block where this transfer was in);
- block_number (block number where this transfer was in);
- block_hash (hash of the block where this transfer was in);

The dataset is hosted on Google Cloud and accessible via Google Big Query.

2) Query of on-chain data: We first scraped all public exchanges' addresses from etherscan.io[5]. Then we used these addresses to query all the transactions coming from any wallet to an exchange address, and all transactions coming from an exchange wallet to any wallet. More details can be found on our appendix A. We then merged the two datasets and dropped the duplicates based on the *transaction_hash* feature, which uniquely identifies a transaction.

3) Off-chain transaction (trading) data: We queried[6] daily data using the API of Crypto Compare. By iterating requests of 2000 candles, we downloaded the available daily price data[7]. We obtained a dataset with the following features:

- time (timestamp identifying the beginning of the trading interval);
- high (highest price in the trading interval);
- low (lowest price in the trading interval);
- open (opening price of the trading interval);
- volumefrom (volume in the base currency, i.e. the ERC-20 token of interest);
- volumeto (volume in the currency it is traded against, i.e. Bitcoin);
- close (timestamp of the block where this transfer was in);
- conversion (block number where this transfer was in);

For each token, we also queried the daily BTC data [8] matching the quotation start date of the token. This was useful to add another feature related to the price of Bitcoin.

B. Preprocessing

We have at hand both inflows to, and outflows from exchanges.

1) Preprocessing:

- We remove all transactions prior to exchange quotation, i.e. all trading intervals whose close price is equal to 0.
- We remove the first fourteen days of trading to avoid outliers often found right after an asset's quotation.
- We remove all rows with infinite values.
- We remove all returns superior to 100 in absolute value.
- We remove all on-chain data when the off-chain counterpart is missing at the provided timestamps.

2) Features:

- We consider as *volume*, the trading volume in the currency pair that the token is being traded in, i.e. Bitcoin.
- On the on-chain data, we create the dummy variable *_FROM_EXCHANGE*, taking value 1 if the given transaction's $from\_address$ is an exchange wallet, else 0. Likewise, we create the dummy variable *_TO_EXCHANGE*, taking value 1 if the given transaction's $to\_address$ is an exchange wallet, else 0.
- We aggregate on-chain transactions by *value* to match the trading interval available in off-chain data, i.e. we match the trading intervals' frequencies. We compute *count* and *sum* variables of the *value* feature.
  - The *count* variable yields the *_ONCHAIN_TRANSACTIONS* feature, i.e. the number of on-chain transactions in the provided time interval.
  - The *sum* variable yields the *_ONCHAIN_VOLUME* feature, i.e. the number of tokens being sent in the provided time interval.
- The *count* of *_TO_EXCHANGE* yields the *_TO_EXCHANGE_TRANSACTIONS* feature, i.e. the number of on-chain transactions directed to an exchange, in the provided time interval.
- The *count* of *_FROM_EXCHANGE* yields the *_FROM_EXCHANGE_TRANSACTIONS* feature, i.e. the number of on-chain transactions originating from an exchange, in the provided time interval.
- We created the feature *_FROM_EXCHANGE_EWMA* computing the exponentially weighted moving average of the *_FROM_EXCHANGE_TRANSACTIONS* variable with a half-life of 10 trading days.
- We compute *_RETURNS* as the log-returns of the tokens, using

$$return_t = \ln(S_t) - \ln(S_{t-1})$$

where S is the closing price and t the concerned observation.

[4]source: Google BigQuery

[5]exchanges_addresses contains information on 190 exchanges and their public Ethereum addresses.

[6]https://min-api.cryptocompare.com/data/v2/ [...] histoday?tsym=BTCfsym={}limit=2000 where the token name is in place of the empty curly brackets

[7]The code of the requests can be found here

[8]https://min-api.cryptocompare.com/data/v2/ [...] histoday?tsym=USDTfsym=BTClimit=2000

- We compute the variable to predict *_VOL* as the exponentially weighted moving average of returns with a half-life of 10 trading days. The resulting series can be observed in figure 7.

We obtain the following design matrix with respect to volatility regression:

- time
- _TO_EXCHANGE_TRANSACTIONS (the number of on-chain transactions directed towards exchange in the given trading interval)
- _FROM_EXCHANGE_TRANSACTIONS (the number of on-chain transactions originating from an exchange in the given trading interval)
- _FROM_EXCHANGE_EWMA
- _ONCHAIN_TRANSACTIONS (the total number of on-chain transactions in the given trading interval)
- _ONCHAIN_VOLUME (the total number of tokens being sent in the given trading interval)
- high
- low
- volume (the *volumeto* variable)
- BTC_high
- BTC_low
- BTC_volume

The variable to predict is:

- _VOL (Exponentially Weighted Moving average of squared returns)

C. Correlation Matrix Analysis

We computed the correlation matrix of the data and saw that the variables with prices of assets such as high, low and close were very correlated. So we only kept the "close" variables for our analysis. A correlation matrix of the final variables is observable in figure 5 for the tokens ZRX and THETA and in figure 6 for the ENJ token.

II. Model Evaluation and Interpretation of Results for ZRX

A. Statistical Significance Analysis

We computed the F-scores for each feature to see which features are statistically significant. The results can be observed in table I for the ZRX token. They show that all variables are significant at 95% confidence interval. The most significant variables are BTC_close and FROM_EXCHANGE_TRANSACTIONS. The results were similar for the other two tokens with all the variables being statistically significant.

p_val	column_name
1.5e-73	BTC_close
1.5e-51	_FROM_EXCHANGE_TRANSACTIONS
4.7e-22	BTC_volume
1.1e-20	close
2.4e-09	_ONCHAIN_TRANSACTIONS
5.8e-07	volume
6.1e-07	_ONCHAIN_VOLUME
6.0e-06	_VOL
5.9e-02	_BTC_RETURNS
2.0e-01	_TO_EXCHANGE_TRANSACTIONS

Table I: P-values of the statistical significance of the variables in the prediction of the exponentially weighted square returns (for the token ZRX). We can observe that the most significant variables are BTC_close, _FROM_EXCHANGE_TRANSACTIONS.

B. Ridge Regression Evaluation

We fitted a Ridge regression model to the data and used a grid-search cross-validation technique to find its best "alpha" parameter. We obtained a best R^2 score of -0.024. This value of close to 0 shows that the model did not find any predictable information in the data. The value of the R^2 can be negative because we do not compute it on the same data that we train the model on. Therefore, if the model is unable to learn any general knowledge it will have an R^2 close to zero and eventually negative. The results are similar also for the two other tokens and the conclusion is the same. No predictability seems to be able to be captured by a ridge Regression model.

C. Linear Regressions with only one variable

As we suspected that the model might be overfitting the data, we computed linear regressions with only one variable for each of the explanatory variables. The results are visible on the table II. They show that the model is able to fit 24% of the variability of the data for the _FROM_EXCHANGE_TRANSACTIONS and for the BTC_close feature in the case of the ZRX token.

R^2	feature
0.32	BTC_close
0.10	BTC_volume
0.00	_BTC_RETURNS
0.00	_TO_EXCHANGE_TRANSACTIONS
0.24	_FROM_EXCHANGE_TRANSACTIONS
0.04	_ONCHAIN_TRANSACTIONS
0.03	_ONCHAIN_VOLUME
0.10	close
0.03	volume

Table II: R^2 of the regressions of each variable alone against the squared returns.

We can observe that the variables BTC_close and _FROM_EXCHANGE_TRANSACTIONS have a notably higher R^2 than the rest. We plotted the graphs of _VOL vs. BTC_close and _FROM_EXCHANGE_TRANSACTIONS in figure 1 for the case of the ZRX token and in figure 2 for the case of the THETA token. Although the plots show

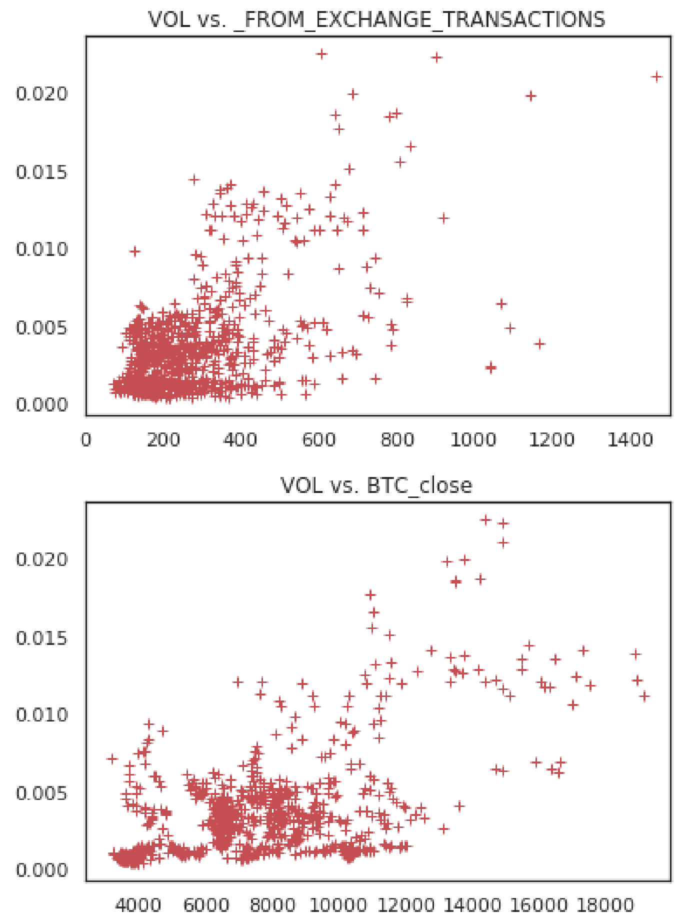


Figure 1: Graph of _VOL vs. BTC_close and _FROM_EXCHANGE_TRANSACTIONS for the ZRX token

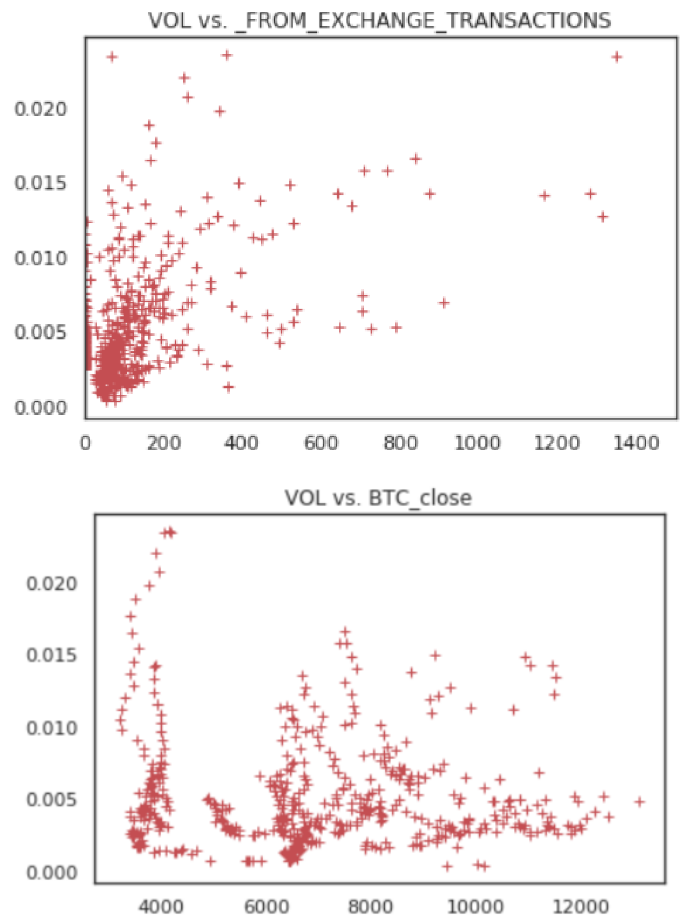


Figure 2: Graph of _VOL vs. BTC_close and _FROM_EXCHANGE_TRANSACTIONS for the THETA token

some structure, as they do not seem to illustrate completely independent random relationships, the noise in the data is visible and no particular relationship is apparent.

D. Random Forest Evaluation and Analysis

We fitted a random forest with the *GridSearchCV* tool of *sklearn*. The hyperparameters that we tuned were 'n_estimators': [5, 10, 20, 40, 80] and 'max_depth':[1,2,3,4, 10]. We did not find a positive R^2 for any of the three tokens. This might be explained by the fact that data is very noisy and the model is not able to predict unseen data.

E. Dynamic regression models

1) Statistical significance of variables: To answer the question of whether our dependent variables are useful in an ARMAX model, we fit two regressive ARMA models. One with all the explanatory variables another with only the volume variable. The results for the ZRX token can be seen in figure 10. We observe in the figures that the only statistically significant coefficients are the AR(1) and the BTC_close ones. The AIC and BIC are lower in the model with all the variables showing that they have some explanatory power. The correspondent regression results for the other tokens can be observed in figures 11 and 12. Although the on-chain variables did not come up as significant for the ZRX model, they are useful in the other two tokens dynamic regressions. Indeed, for the THETA coin we observe that the significant variables are AR(1), MA(1), _TO_EXCHANGE_TRANSACTIONS, _FROM_EXCHANGE_TRANSACTIONS, _BTC_close and _volume. And for the ENJ coin we observe that the significant variables are AR(1), MA(1), _FROM_EXCHANGE_TRANSACTIONS, _BTC_close and _volume.

2) Analysis of the magnitude and sign of the coefficients: The results show that the relationship between an increase in variance and Bitcoin price and transactions towards exchanges is positive for some tokens: For the ZRX token the BTC_close variable is significant. It has a beta coefficient of 0.02. For THETA and ENJ, the coefficients are 0.05 and 0.03 respectively. The fact that the data were normalized before doing the regression allows the beta coefficients to be compared.

We can compare the magnitude of the beta coefficient of the BTC_close variable (Bitcoin price) with the beta coefficient of the trading volume (volume variable). This variable shows the market trading volume. The market volume beta coefficient is 0.05 in the ARIMAX model with all the features for the ZRX token. For the THETA token this coefficient is 0.15, and for the ENJ token 0.21. This shows that the market volume variable is the most informative one when it comes to volatility changes.

The results show that there is sometimes a relationship between on-chain transactions and volatility. The _TO_EXCHANGE_TRANSACTIONS variable is significant for the THETA coin. It has an estimate of 0.068. We can interpret this result as the fact that a higher volatility is expected when more transactions towards exchanges' wallets are recorded. This result seems reasonable. The _FROM_EXCHANGE_TRANSACTIONS variable is significant for THETA and ENJ. It has an estimate of -0.068 (THETA) and -0.12 (ENJ). We can interpret this result as the fact that a lower volatility is expected when more transactions from exchanges wallets are recorded. This result seems reasonable since outflows of money from exchanges could result in fewer trade volume.

F. Why do we get statistical significance of the data series in the ARIMA models and no R^2 in the regressions?

We believe that the reasons for which the Linear Regression and Random Forest models are unable to predict out of sample data are because they try to predict the volatility directly from the features at one point in time. On the other hand, the ARIMAX models aggregate information from a point in time and a series of past events. To make the comparison more fair we should predict changes in volatility but giving to the model the information about what volatility has been up to time t.

CONCLUSION

We have seen that all variables are statistically significant in a preliminary F test. Then we have fitted a Ridge Regression and a Random Forest models but we were unable to predict any significant variance of the data. By doing a Linear regression with only one variable at a time, we have determined that the two most meaningful variables were BTC_close and _FROM_EXCHANGE_TRANSACTIONS. The training data R^2 were high among those variables (around 25%), but the graphs show that the data are very noisy, and the relationships are not clear. Finally we have tested the statistical significance of our features in a comparison between ARMAX models. They indicate, for the ZRX token, that the only two useful features are the trading volume and Bitcoin price BTC_close. These two variables improve the BIC and AIC of the models. For THETA and ENJ the _TO_EXCHANGE_TRANSACTIONS and _FROM_EXCHANGE_TRANSACTIONS are significant. And the sign of their coefficients seems consistent with what would be expected.

In conclusion, the trading volume variable seems to be the most informative feature for volatility. The transactions on-chain with exchanges seem to carry some information, but they need to be incorporated in dynamic regression models as they are not able to predict volatility on their own due to the noise present in financial market data.

REFERENCES

[1] J. M. Griffin and A. Shams, "Is Bitcoin Really Un-Tethered?", en, Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3195066, Oct. 2019. [Online]. Available: https://papers.ssrn.com/abstract=3195066 (visited on 11/17/2019).

[2] F. BLACK, "Noise", *The Journal of Finance*, vol. 41, no. 3, pp. 528–543, 1986. DOI: 10.1111/j.1540-6261.1986.tb04513.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1986.tb04513.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1986.tb04513.x.

APPENDIX

A. Querying the on-chain data

We used the following SQL query to query on-chain data transactions directed TO exchanges:

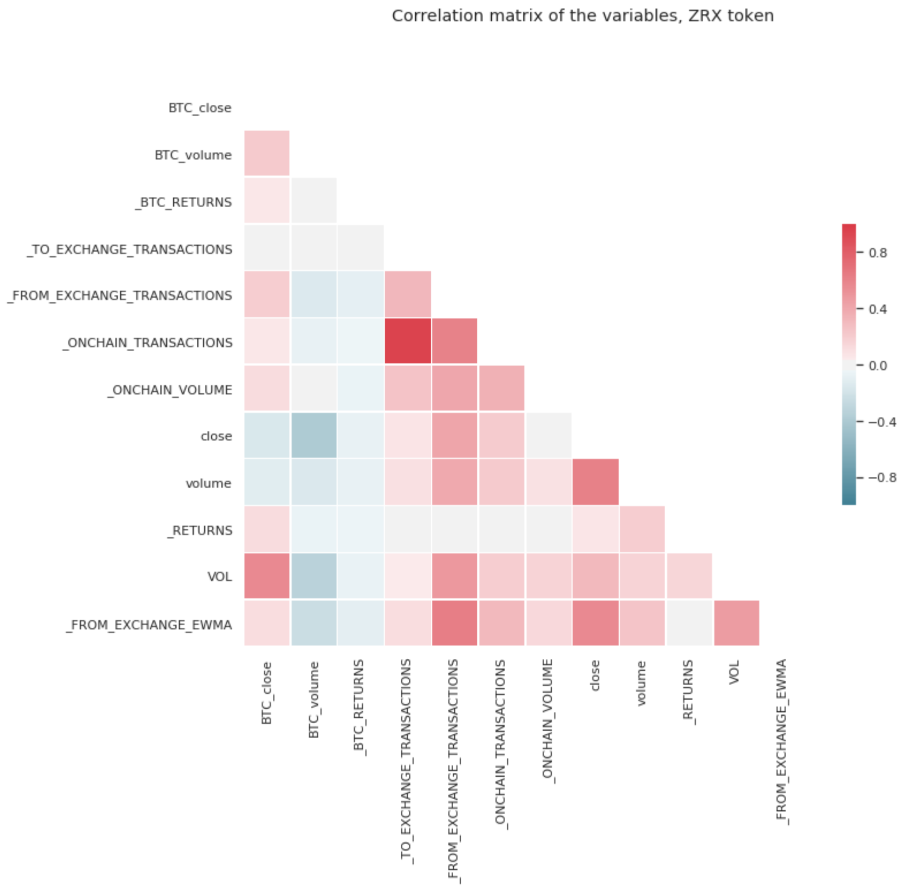
```
SELECT
  *
FROM
  `bigquery-public-data.crypto_ethereum\
  token_transfers` AS transfers
WHERE to_address IN
(exchanges_addresses)
AND token_address=\
'0x3883f5e181fccaf8410fa61e12b59bad963fb645'
```

Figure 3: SQL query to get all transactions involving THETA which were directed TO identified exchanges' wallets. In this example, token_address is THETA's smart contract address.

We then queried on-chain data transactions to any wallet FROM exchanges, using the following query:

```
SELECT
  *
FROM
  `bigquery-public-data.crypto_ethereum\
  token_transfers` AS transfers
WHERE from_address IN
(exchanges_addresses)
AND token_address=\
'0x3883f5e181fccaf8410fa61e12b59bad963fb645'
```

Figure 4: SQL query to get all transactions involving THETA which were coming FROM identified exchanges' wallets. In this example, token_address is THETA's smart contract address.



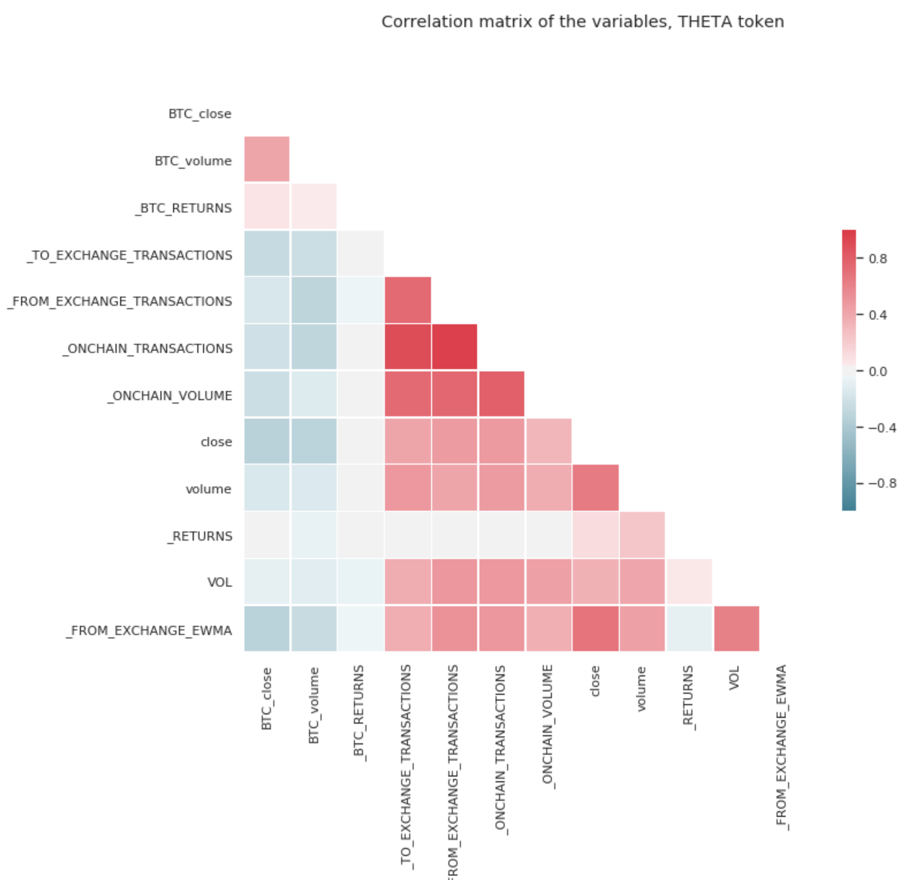


Figure 5: Correlation Heatmap of the variables of the tokens ENJ and THETA. We can observe that the _RETURNS variable is not significantly correlated with the rest of variables. The _VOL variable is significantly correlated with BTC_close and _FROM_EXCHANGE_TRANSACTIONS for ZRX. For THETA, however, the _VOL variable is not correlated greatly with any feature although it presents more dependence that the _RETURNS variable.

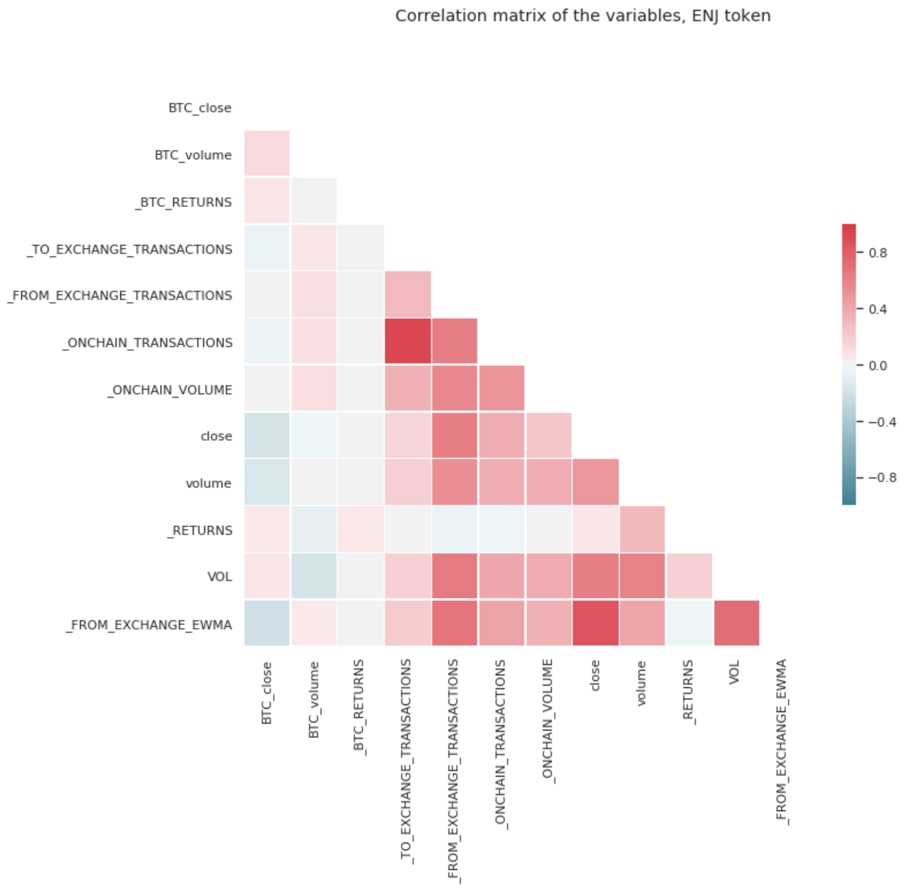


Figure 6: Correlation Heatmap of the variables for the ENJ token.

B. Integration of the _VOL process

The _VOL process is clearly integrated as we can see on figure 8. For this reason we use an ARIMA model with the first difference of _VOL I=1. In figure 9, we can observe that the _VOL process in not integrated for the THETA token.

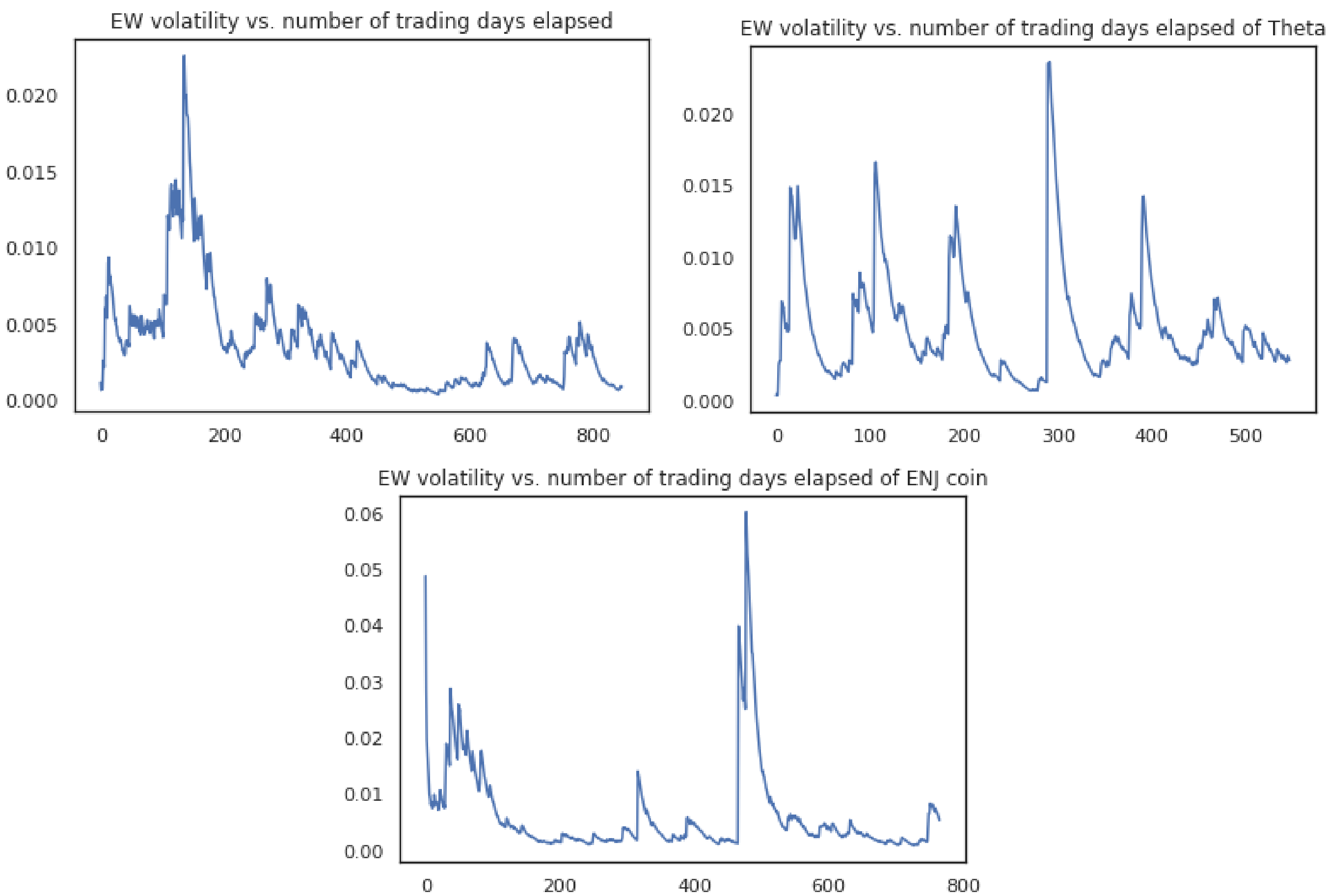


Figure 7: Plot of EW moving average of square daily returns of ZRX (left), THETA (right), ENJ (down).

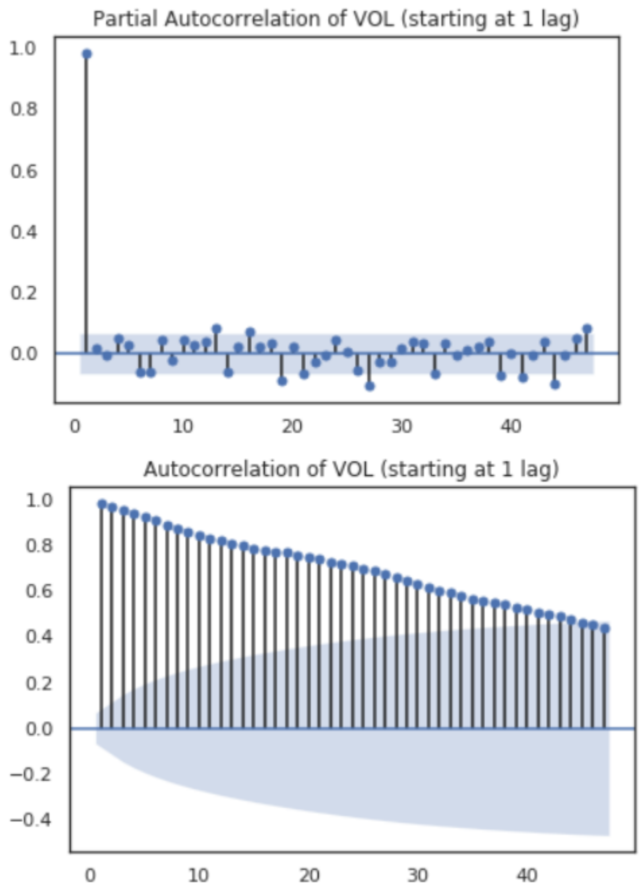


Figure 8: The Pacf and Acf of the EWMA Volatility process _VOL.

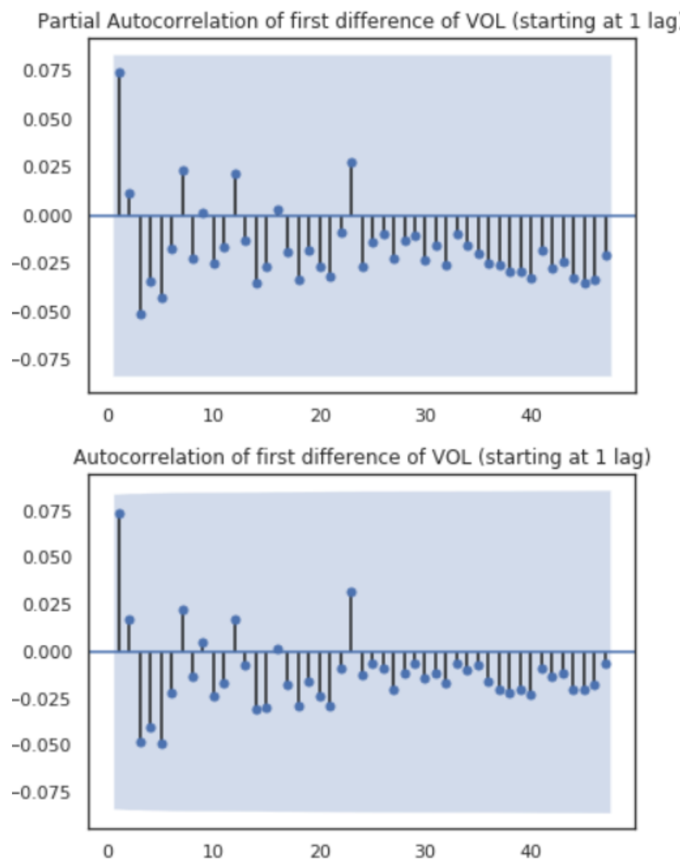


Figure 9: ACF and PACF for the fist difference of the Theta token.

```
Normal ARIMAX(1,1,0)
======================================================= ==================================================
Dependent Variable: Differenced VOL                     Method: MLE
Start Date: 1                                           Log Likelihood: -169.6141
End Date: 546                                           AIC: 347.2281
Number of observations: 545                             BIC: 364.4313
==========================================================================================================
Latent Variable                          Estimate   Std Error  z        P>|z|    95% C.I.
======================================== ========== ========== ======== ======== =========================
AR(1)                                    0.0463     0.0407     1.1387   0.2548   (-0.0334 | 0.1261)
Beta 1                                   0.001      0.0141     0.0712   0.9433   (-0.0267 | 0.0287)
Beta volume                              0.1103     0.0142     7.7734   0.0      (0.0825 | 0.1381)
Normal Scale                             0.3303
==========================================================================================================
```

```
Normal ARIMAX(1,1,1)
======================================================= ==================================================
Dependent Variable: Differenced VOL                     Method: MLE
Start Date: 1                                           Log Likelihood: 253.2723
End Date: 847                                           AIC: -488.5445
Number of observations: 846                             BIC: -445.8799
==========================================================================================================
Latent Variable                          Estimate   Std Error  z        P>|z|    95% C.I.
======================================== ========== ========== ======== ======== =========================
AR(1)                                    -0.5129    0.1122     -4.5731  0.0      (-0.7328 | -0.2931)
MA(1)                                    0.4386     0.1158     3.7875   0.0002   (0.2116 | 0.6656)
Beta 1                                   -0.0       0.0089     -0.0017  0.9986   (-0.0174 | 0.0174)
Beta _TO_EXCHANGE_TRANSACTIONS           -0.0003    0.0063     -0.0553  0.9559   (-0.0128 | 0.0121)
Beta _FROM_EXCHANGE_TRANSACTIONS         -0.0136    0.0097     -1.4038  0.1604   (-0.0325 | 0.0054)
Beta BTC_close                           0.0249     0.0091     2.7366   0.0062   (0.0071 | 0.0427)
Beta volume                              0.0524     0.0078     6.7072   0.0      (0.0371 | 0.0677)
Beta _BTC_RETURNS                        0.0        0.0055     0.0058   0.9953   (-0.0108 | 0.0109)
Normal Scale                             0.1794
==========================================================================================================
```

```
Normal ARIMAX(1,1,0)
======================================================= ==================================================
Dependent Variable: Differenced VOL                     Method: MLE
Start Date: 1                                           Log Likelihood: 242.6462
End Date: 847                                           AIC: -475.2923
Number of observations: 846                             BIC: -451.5897
==========================================================================================================
Latent Variable                          Estimate   Std Error  z        P>|z|    95% C.I.
======================================== ========== ========== ======== ======== =========================
AR(1)                                    -0.0746    0.0351     -2.1267  0.0334   (-0.1433 | -0.0058)
Beta 1                                   0.0        0.0062     0.0019   0.9984   (-0.0122 | 0.0123)
Beta BTC_close                           0.015      0.0063     2.3824   0.0172   (0.0027 | 0.0274)
Beta volume                              0.0319     0.0065     4.9013   0.0      (0.0191 | 0.0446)
Normal Scale                             0.1816
==========================================================================================================
```

Figure 10: Results of ARIMAX models for ZRX using the most explanatory variables (middle), using only the volume variable to predict square returns (up) and using only the statistically significant coefficients (down). The variables with statistically significant p-values are AR(1), BTC_close, and volume

```
Normal ARIMAX(1,1,1)
======================================================= ==================================================
Dependent Variable: Differenced VOL                     Method: MLE
Start Date: 1                                           Log Likelihood: -150.4435
End Date: 546                                           AIC: 318.887
Number of observations: 545                             BIC: 357.5941
==========================================================================================================
Latent Variable                          Estimate   Std Error  z        P>|z|    95% C.I.
======================================== ========== ========== ======== ======== =========================
AR(1)                                    -0.3679    0.0875     -4.2058  0.0      (-0.5393 | -0.1964)
MA(1)                                    0.4432     0.077      5.7558   0.0      (0.2923 | 0.5941)
Beta 1                                   0.0015     0.0197     0.0745   0.9406   (-0.0371 | 0.0401)
Beta _TO_EXCHANGE_TRANSACTIONS           0.0686     0.0252     2.722    0.0065   (0.0192 | 0.118)
Beta _FROM_EXCHANGE_TRANSACTIONS         -0.0633    0.0251     -2.5225  0.0117   (-0.1125 | -0.0141)
Beta BTC_close                           0.0451     0.0205     2.2041   0.0275   (0.005 | 0.0852)
Beta volume                              0.1505     0.0174     8.6326   0.0      (0.1163 | 0.1846)
Beta _BTC_RETURNS                        0.0018     0.0119     0.1518   0.8793   (-0.0215 | 0.0252)
Normal Scale                             0.3189
==========================================================================================================
```

Figure 11: Results of ARIMAX models for THETA using the most explanatory variables. The variables with statistically significant p-values are AR(1), MA(1), _TO_EXCHANGE_TRANSACTIONS, _FROM_EXCHANGE_TRANSACTIONS, BTC_close, and volume.

```
Normal ARIMAX(1,1,1)
======================================================= ==================================================
Dependent Variable: Differenced VOL                     Method: MLE
Start Date: 1                                           Log Likelihood: 62.5871
End Date: 765                                           AIC: -107.1741
Number of observations: 764                             BIC: -65.427
==========================================================================================================
Latent Variable                          Estimate   Std Error  z        P>|z|    95% C.I.
======================================== ========== ========== ======== ======== =========================
AR(1)                                    -0.3004    0.0507     -5.9218  0.0      (-0.3998 | -0.201)
MA(1)                                    0.2087     0.0532     3.9263   0.0001   (0.1045 | 0.3129)
Beta 1                                   -0.0065    0.0098     -0.6661  0.5053   (-0.0256 | 0.0126)
Beta _TO_EXCHANGE_TRANSACTIONS           -0.0038    0.0083     -0.4614  0.6445   (-0.0201 | 0.0125)
Beta _FROM_EXCHANGE_TRANSACTIONS         -0.1165    0.0109     -10.7226 0.0      (-0.1378 | -0.0952)
Beta BTC_close                           0.033      0.0099     3.319    0.0009   (0.0135 | 0.0525)
Beta volume                              0.2105     0.0117     17.9845  0.0      (0.1875 | 0.2334)
Beta _BTC_RETURNS                        0.0023     0.008      0.2877   0.7736   (-0.0133 | 0.0179)
Normal Scale                             0.2229
==========================================================================================================
```

Figure 12: Results of ARIMAX models for ENJ using the most explanatory variables. The variables with statistically significant p-values are AR(1), MA(1), _FROM_EXCHANGE_TRANSACTIONS, BTC_close, and volume