Massachusetts Institute of Technology

6.419 Statistics, Computation and Applications

Final Project Report

Undergraduate Finance 2 Group

Cryptocurrency Market Analysis

Authors

Yuhao Chen, Andrew Turner, Andy Tso, Sean Ko, Taras Zhylenko

December 30, 2018

1 Introduction

The cryptocurrency market is an interesting new part of the financial world, with the advent of blockchain technology showing great promise for the future of decentralized systems. However, the cryptocurrency market is not well understood, as people question the inherent value of cryptocurrencies as well as the legitimacy of cryptocurrency exchanges (e.g. risk of market manipulation). In order to get a good understanding of how the cryptocurrency markets work, we attempt to answer the following questions:

1. What currencies serve as a good representation of the whole market?
2. How to quantify the goodness of such representation?
3. How we can predict the price movement efficiently?
4. How to detect when the market is stable and what are the consequences of it being unstable?
5. How is the behavior of the market on the hour horizon is different from its behavior on the week horizon?

In order to investigate the first two questions, we will employ network analysis that will show us how different currencies are related in terms of trade activity and correlation of their price movements. We also find network analysis to be beneficial, as it allows us to capture interdependencies and mutual positioning of the cryptocurrencies in the market (one of our trivial findings show that most currencies are exchanged via bitcoin). We proceed with analyzing time series with classical techniques and then move to an RNN model. Finally, we compare our findings on the macro scale with the intraday analysis, which attempts to look at trading activity on a smaller time scale.

2 Important Currencies Through Time

2.1 Questions

The cryptocurrency market is composed of hundreds of individual currencies, and as a preliminary analysis, we attempt to find which currencies are important and which are the most active avenues of trade 2016-2018. We choose 2016 as a starting point to conduct analysis, as we deemed this to be the point at which the cryptocurrency market started to take off. Furthermore, we explore which interactions are crucial in the economy and its evolution through time.

2.2 Methods

We first obtain our handful of important currencies by ranking the coins by market share. Due to the limited availability of exchange volume in our dataset, we choose the top 6 among our available currencies: Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC), Monero (XMR), Dash (DASH), and ZCash (ZEC), currently capturing a significant 60% of the total cryptocurrency market cap. The cutoff was chosen at ZCash due to decaying market cap of coins by marketcap beyond it. That is, adding the next largest coin adds less than 1% more of the total market cap.

For each of the years $2016, 2017, 2018$, we then build a complete graph $G = (V, E)$ where V is given by the set of the currencies under consideration, and define an activity function over edges:

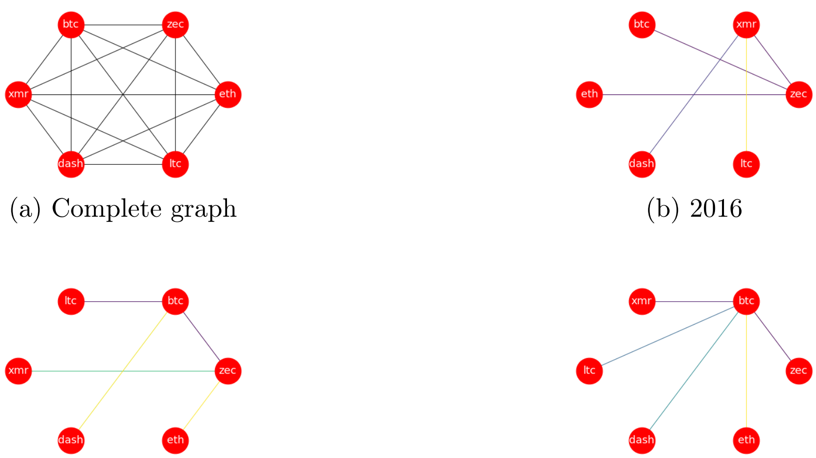
$$\text{activity}(u, v) = \frac{\text{dollars traded}(u, v)^2}{\text{marketcap}(u) \cdot \text{marketcap}(v)}$$

The choice is given as such since we desire to capture a normalized quantity of trade, such that the activities among edges do not differ large orders of magnitude. For a given year, we consider the total dollars traded between the two currencies, and the market cap of a given year is given by the average over daily market caps. We interpret the activity of an edge to be a measure of the extent to which a pair of currencies is traded with respect to how much trade is possible between the two currencies.

We proceed to find the maximum weight spanning tree, or equivalently the minimum weight spanning tree of the inverse activty graph if we wish to think in terms of a measure of closeness/distance.

2.3 Results

Below we plot the MST's obtained. We observe only after the crash in early 2018 does the trade activity heavily centralize on Bitcoin. Although Bitcoin had been dominant, our measure of activity shows that smaller currencies had their markets stretched further to their limits, by trading indirectly around Bitcoin.



(a) Complete graph

(b) 2016

(c) 2017

(d) 2018

Figure 1: Activity MST's

3 Evolution of Correlations Between Main Currencies

3.1 Questions

Apart from knowing the market trade volume correlations, we want to dig more into the price series correlations. We want to answer some questions: What are the correlations of cryptocurrency prices? How do the correlations changes over time? Can we find some interesting patterns by looking at their correlations? We chose 2016 as a starting point because after this year the data is ample so we can avoid some bias.

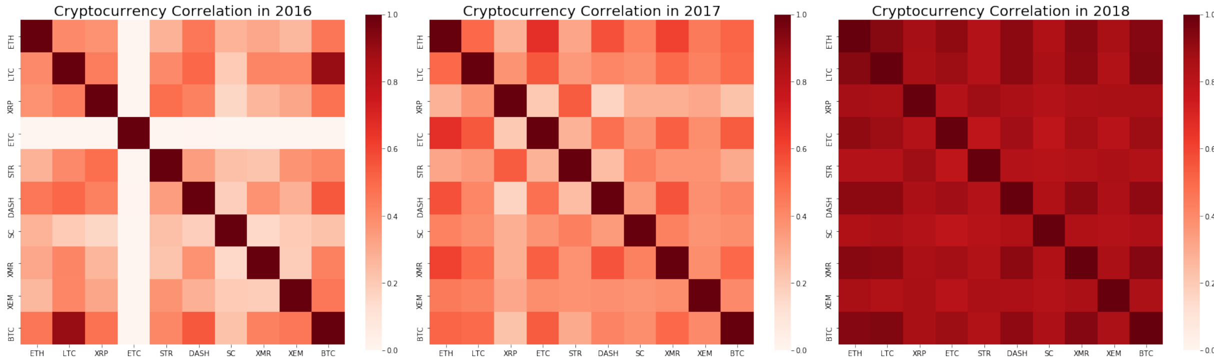


Figure 2: Heat Map of Correlation Evolution

3.2 Methods

We conducted time series analysis and network analysis in this part. We obtained 28 cryptocurrencies daily price from 2016 to 2018, which can almost cover the whole market capital. We computed the time series returns:

$$\text{Return}_t = \frac{\text{Price}_t - \text{Price}_{t-1}}{\text{Price}_{t-1}}$$

In 28 return series, we computed the pairwise pearson correlation coefficient and add them to matrix in 2016, 2017 and 2018. So we have three correlation matrix for 2016, 2017 and 2018, respectively. After that, we plotted the heat map to visualize the evolution of correlation over the three years, which can be seen in results.

In order to find some patterns in the correlations, we determined to use network analysis to achieve this goal. The nodes in the network represent cryptocurrencies and the edges represent the correlations between edges. We don't want to the graph are fully connected so we conducted the ADF hypothesis test(Augmented Dickey–Fuller Test) to determine whether the correlation is significant. The null hypothesis of this test is equal to that the difference of the two series are random walk, so if we can reject the null hypothesis we can say they are cointegrated and have a significant correlation. First, for each two price time series, we unified them. Then we used ADF test to pairwise time series and get the p-value. If the p-value is lower than 5%, the edge will exist in the graph, which means they have a significant correlation. The weight of existing edges will be the correlation coefficient.

3.3 Results

Three heat maps (see figure 2) show the evolution of correlation from 2016 to 2018. We choose 9 out of the 28 cryptocurrencies to show the heat map more clearly. In 2016, we can see that they are not well correlated with each other, but 2017's heat map is darker which means their correlation increased in 2017. In 2018, the heat map converges to almost complete correlation. Thus, we can draw a conclusion that the cryptocurrency market is becoming more correlated as time progresses. This is in the wake of the market crashing and continuing to fall - a similar level of correlation was present in the stock market when it crashed in 2008.

After that, we can look at the graph we built and found that the currency USDT (Tether) is especially interesting (see figure 3). It is in the center of the graph and has the highest weighted degree centrality, which means it has the most significant correlation with other currencies. We plot the price history of USDT, BTC (Bitcoin) and ETH (Ethereum) to verify our assumption (see figure 4). It shows that USDT precisely captures the spike of BTC and ETH as well as the slump from both. So we can draw a conclusion that USDT is a great indicator of the market. Even a small change in the USDT price (from 1 dollar to 1.08 dollar) can reflect significant changes in the market.

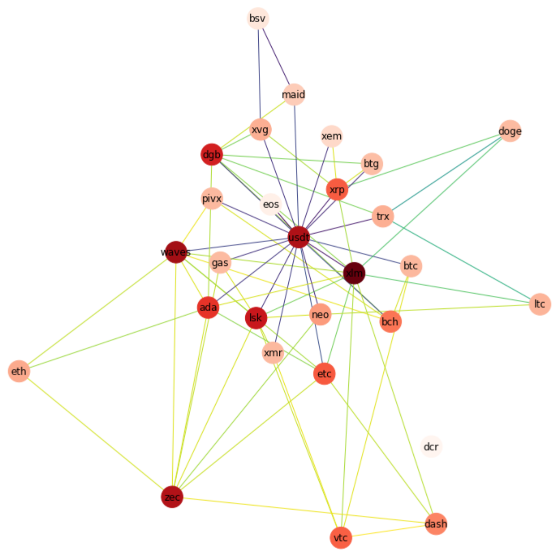


Figure 3: Correlation Network

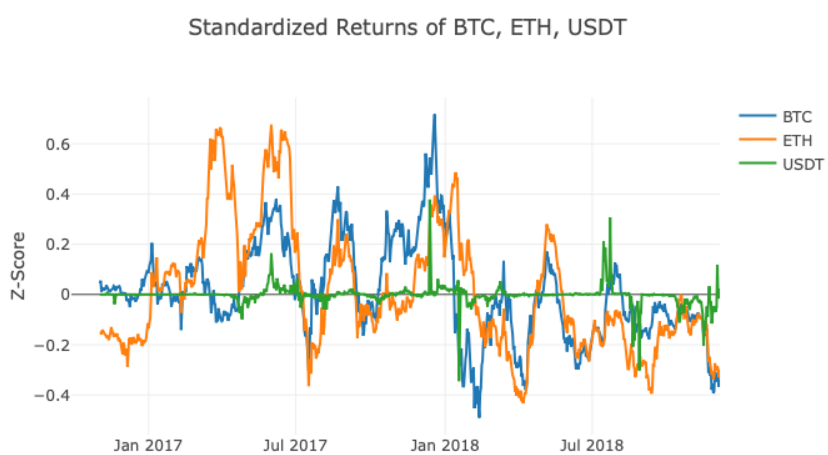


Figure 4: USDT, BTC and ETH historical price

4 Price Series and Market Regimes

4.1 Motivation for analysis

Next, we decided to analyze the cryptocurrency market as characterized by time series of prices of different coins. Consistent with our previous observations and exploratory analysis, we observe that the market was quite unstable in the period starting from September 2017 until June 2018 (market crash). Therefore, we first attempt to fit the simplest models such as AR, MA, ARMA and multivariate Gaussian Process to the daily data available prior to the market crash (May 2016 through October 2017), as we suspect the market to be governed by some stationary random process. However, we were not able to fit any of these models successfully. Our algorithm for fitting ARMA did not converge and Gaussian Process was predicting values way off the true observations. After taking a second look at the data, we came to understand that the data we are dealing with (namely, daily price changes) cannot be assumed to have the same distribution throughout the time frame considered. For example, the series exhibits clear upward trend starting at May 2017. Therefore, it is not helpful to use any model fitted on data collected before May 2017 to predict the price change behavior after it and vice versa.

Starting from this issue, we decide to figure out the location and duration of all relatively stable and relatively chaotic periods. We will use this information to train our future models on a reliable time frame and also, to assess market efficiency during the times of instability.

4.2 Methods

In order to capture behavior of the market, especially, in second half of the considered period when most of the cryptocurrencies are very correlated with one another (see discussion above), we decided to base our analysis on the 3 coins with the biggest market cap: Bitcoin (BTC), Ethereum (ETH) and Ripple (XRP). Our logic is the following: these three currencies are represented by a well-correlated market, and any market signals can be appropriately measured by the reaction of BTC, ETH an XRP. We are looking for the periods of instability in the market and we will attempt to find them using Kolmogorov-Smirnov Test. For every date in a time frame of interest, we take 20 dates prior to the one taken and 20 more after that and apply K-S test to the price changes as inferred from these 20-day periods. We do this for price series of each coin and conclude market instability if K-S indicated rejection of null hypothesis (both samples coming from the same distribution) for all three coins. In order to assess market efficiency, we will be running a 40-day momentum investment strategy on a portfolio of three coins above, weighted equally.

4.3 Results

First, we display the results of our KS-testing (see figure 5). We see that the regime change points (where KS rejected null hypothesis for all three coins) capture oscillations of the market very well (See February 2017 - June 2017 and October 2017 - January 2018 periods). It means that fitting the models on any training set that contains marked intervals will lead to erroneous predictions on the testing set as we would essentially be training our model on the outliers. Also, we see that the periods when the momentum strategy yields the highest returns coincide with regime change regions that we have found. It reinforces the assumption that in times of uncertainty, market offers good investing opportunities (see figure 5).

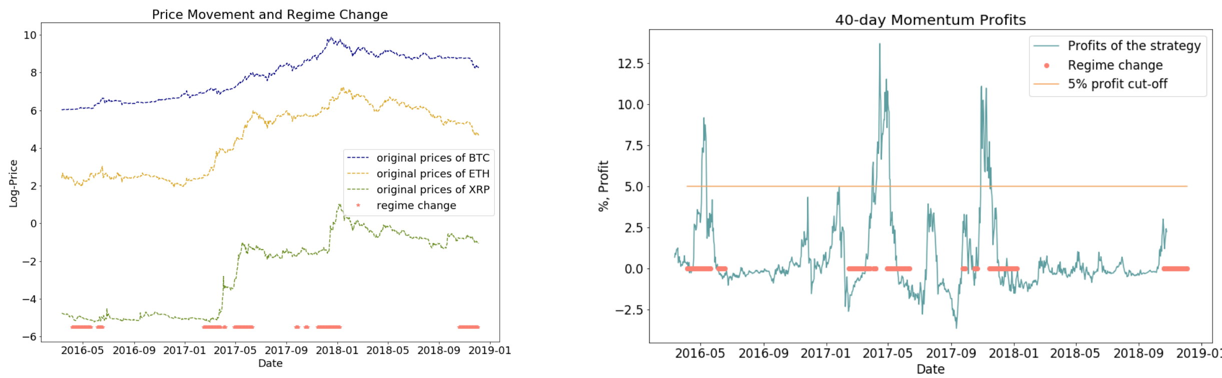


Figure 5

5 Predicting Bitcoin Prices with Recurrent Neural Networks

5.1 Questions

The cryptocurrency market attracts people to trade these currencies to make money based on the difference in prices when bought and sold. If people could accurately predict the prices of Bitcoin, then people could be making accurate decisions on when to buy and sell to increase their returns. With time series data, we apply a simple machine learning model: we use a recurrent neural network and feed in past bitcoin prices. We will take a look into this method.

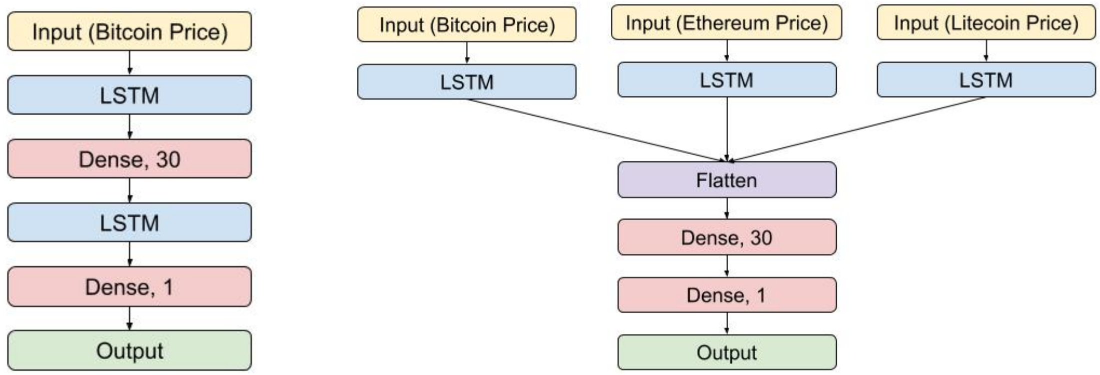
5.2 Methods

We first obtain historic prices of Bitcoin for the last five years so that we have a large enough dataset to use the last 30 days in Bitcoin prices to predict tomorrow's prices. Then we construct our first baseline RNN architecture as shown in Figure 6a. We'll feed the 30 days of bitcoin prices into a first LSTM block, whose outputs get fed into a dense layer with 30 nodes, followed by a second LSTM block, connected a dense layer with one final output neuron.

For our second model, we will also obtain the historic prices of two other large cryptocurrencies - Ethereum and Litecoin - for the last five years as well. We modify our first RNN architecture slightly, as shown in Figure 6b to include separate LSTM blocks for each of these set of inputs. We then concatenate the outputs of these three LSTM blocks and feed it through a dense layer with 30 nodes and then to one final node.

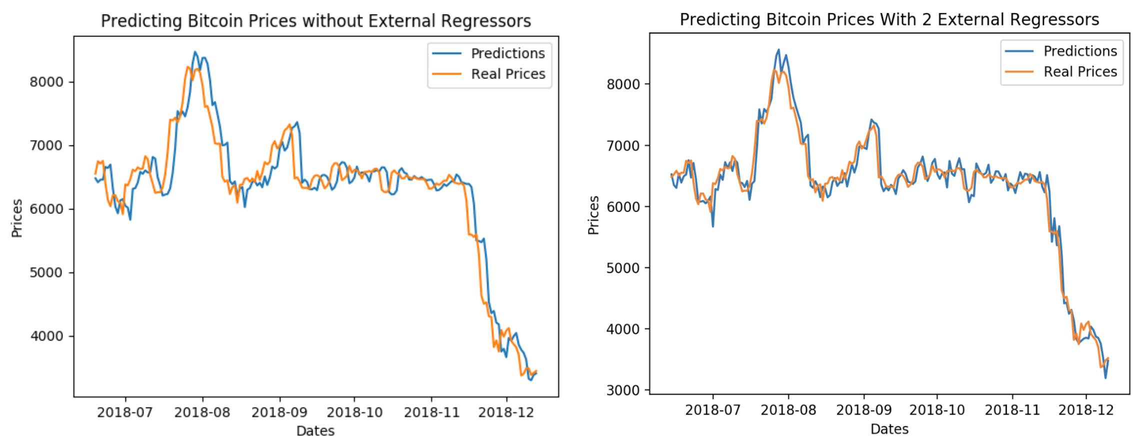
5.3 Results

After training and testing both models, we obtain the prediction results shown in Figure 7a for the baseline RNN model and in Figure 7b for the RNN with external regressors. From this, we can immediately see the striking lag in predictions from the baseline RNN. This martingale-like model has essentially just used the previous day's value as the best prediction for tomorrow's prices. To combat this effect, we have added the external regressors of other coin prices. We can see a better fit to the Bitcoin prices, but there is still a slight lag which could potentially be because Bitcoin prices are still a regressor in this model.



(a) (b)

Figure 6: Architectures of the baseline RNN (a) and the RNN with external regressors (b)



(a) (b)

Figure 7: Prediction results for the baseline RNN model (a) and the RNN with external regressors (b)

6 Intraday Analysis

6.1 Questions

Traditional financial markets are huge and very efficient. For this reason, inefficiencies (i.e. arbitrage opportunities) only tend to exist for milliseconds at a time. The most blatant type of inefficiency is that of exchange rates — if in an FX market, one is able to start with a dollar, make a series of trades for other currencies, and finally end up with more than one dollar. We call this an *arbitrage cycle*. In addition, traditional financial markets have many regulations to prevent market manipulation. For this section, we were able to obtain a free trial for an API that provides data at a frequency of 10 minutes, so that we could attempt to answer the following questions:

1. How inefficient is the cryptocurrency market?
2. How much market manipulation exists in the cryptocurrency market?
3. How is inefficiency and market manipulation reflected in relative value fluctuations and *flows* between cryptocurrencies?

Intraday data is very expensive — the best that we could do for our analysis was obtain ten-minute frequency data for 8 major cryptocurrencies for a two week period in June, 2018. These two weeks were well after the market crash in January of 2018, and so it is a more stable time to analyze these questions (which relate more to the market itself than to market events). Clearly, a two week period is not enough to draw final conclusions, but we believe that we have a good start. In addition, we have a solid prior on the question of market manipulation: it is well known that "pump-and-dump" schemes happen regularly. There are even large communities that plan to "pump" the price of a currency and then subsequently sell very quickly.

6.2 Data & Metrics

We were able to obtain our data for exchnage rates between the following cryptos (and USD) for June 1, 2018 through June 14, 2018 from `coinapi.io`:

1. Bitcoin (BTC); June 2018 Market Cap: \$ 112 B
2. Ethereum (ETH); June 2018 Market Cap: \$ 50 B
3. Ripple (XRP); June 2018 Market Cap: \$ 17 B
4. Litecoin (LTC); June 2018 Market Cap: \$ 5 B
5. Bitcoin Cash (BCH); June 2018 Market Cap: \$ 15 B
6. Ethereum Classic (ETC); June 2018 Market Cap: \$ 1.5 B
7. Tether (USDT); June 2018 Market Cap: \$ 2.5 B

Our data had the following features for each coin: Next, we added columns of percent change over

	price_close	price_high	price_low	price_open	time_period_end	time_period_start	trades_count	volume_traded
0	0.077339	0.077420	0.077171	0.077218	2018-06-01T00:10:00.0000000Z	2018-06-01T00:00:00.0000000Z	1426	1145.538
1	0.077446	0.077455	0.077250	0.077346	2018-06-01T00:20:00.0000000Z	2018-06-01T00:10:00.0000000Z	1466	1142.853
2	0.077447	0.077500	0.077364	0.077442	2018-06-01T00:30:00.0000000Z	2018-06-01T00:20:00.0000000Z	1089	751.154
3	0.077448	0.077500	0.077375	0.077447	2018-06-01T00:40:00.0000000Z	2018-06-01T00:30:00.0000000Z	963	532.919
4	0.077413	0.077500	0.077320	0.077386	2018-06-01T00:50:00.0000000Z	2018-06-01T00:40:00.0000000Z	903	555.771

Figure 8: OHCLV Data from Coin API

the ten minute windows, along with an associated momentum (flow) metric:

$$\text{change} = \frac{\text{price close - price price open}}{\text{price open}}, \qquad \text{momentum} = \text{change} \cdot \text{volume}.$$

We proceeded to standardize these variables for each coin independently, in order to analyze the relative value fluctuations among cryptocurrencies.

6.3 Methods

Our general approach to analyzing the exchange rates was to construct an eight node network (each node being one of the seven cryptos or USD), with edges weighted by exchange rate. Then we used the algorithm descibed below to find arbitrage cycles at each point in time.

In addition, we constructed a flow network with the same nodes, and with directed edges weighted by absolute value of momentum, where the direction pointed to the crypto which was rising in value w.r.t. the other one.

6.3.1 Arbitrage Cycle Algorithm

At each point in time, we run the following algorithm on the exchange rate matrix, which has the form:

	BTC	ETH	BCH	ETC	LTC	XRP	USDT	USD
BTC	0.000000	12.955717	7.516988	487.567040	62.849601	12207.031250	7444.04000	7430.00
ETH	0.077186	0.000000	0.578925	37.647767	4.853662	941.397976	574.07000	573.45
BCH	0.133032	1.727340	0.000000	0.000000	0.000000	0.000000	990.41000	986.00
ETC	0.002051	0.026562	0.000000	0.000000	0.000000	0.000000	0.00000	0.00
LTC	0.015911	0.206030	0.000000	0.000000	0.000000	0.000000	118.47000	118.41
XRP	0.000082	0.001062	0.000000	0.000000	0.000000	0.000000	0.60984	0.00
USDT	0.000134	0.001742	0.001010	0.000000	0.008441	1.639774	0.00000	0.00
USD	0.000135	0.001744	0.001014	0.000000	0.008445	0.000000	0.00000	0.00

Figure 9: Exchange Rate Matrix

For the detection of arbitrage cycles in our graph, we consider a modified Floyd-Warshall algorithm. We use Floyd-Warshall, since the only algorithm with a better runtime is Bellman-Ford with time complexity of $O(|V||E|)$. However we have $|E| = O(|V|^2)$ yeilding an equivalent runtime, so we choose Floyd-Warhsall for ease of implementation. The input of the algorithm is an exchange matrix M where M_{ij} is the exchange rate of cryptocurrency i for cryptocurrency j. While originally Floyd-Warshall algorithm computes all-pairs shortest paths, we modify it to consider the problem of computing all-pairs best arbitrage path. Concretely, we may consider a node k to be an intermediate which improves the directed path from node i to node j if it satisfies the condition that $M_{ik} \cdot M_{kj} > M_{ij}$, signifying an exchange sequence of greater profit. The output is a matrix O where entry O_{ij} specifies the maximum amount of currency j which we may obtain starting with 1 unit of currency i. Then we restrict our attention to diagonal elements O_{ii}, checking if any are greater than 1. We give the algorithm in full below:

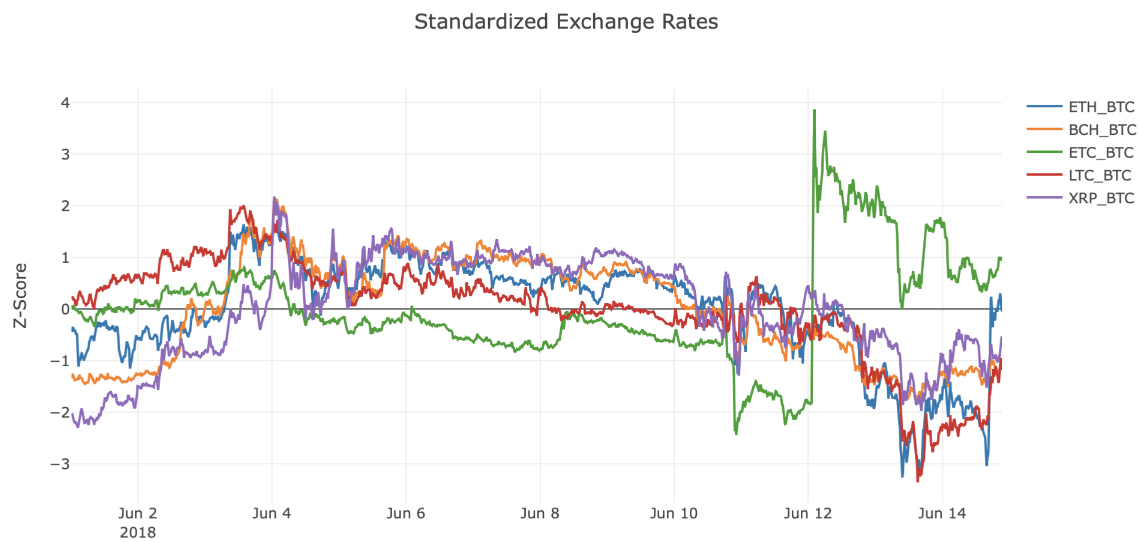
Algorithm 1 Modified Floyd-Warshall

procedure DetectArbitrageCycles(M)
    for k in nodes **do**
        for i in nodes **do**
            for j in nodes **do**
                if $M[i][k] * M[k][j] > M[i][][j]$ **then**
                    $M[i][j] \leftarrow M[i][k] * M[k][j]$
    return M

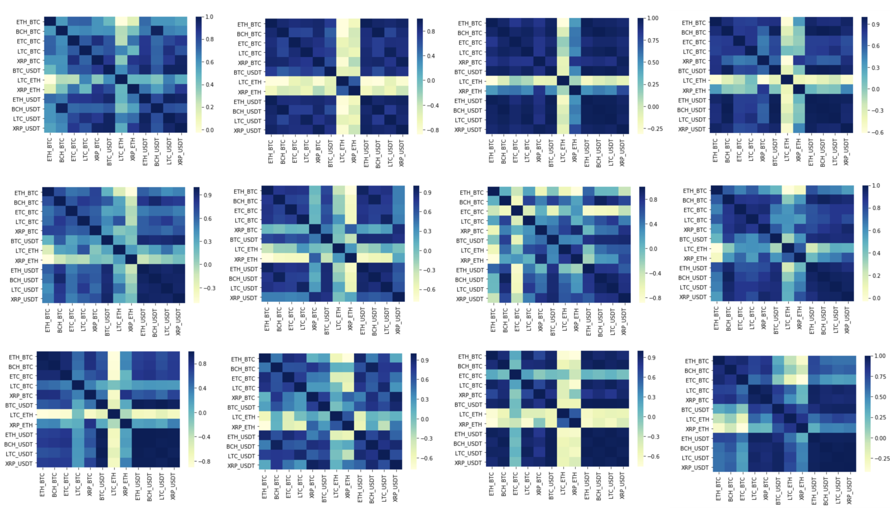
Upon detection of such a cycle, we then use a brute force algorithm to determine such a cycle and its arbitrage opportunity.

6.4 Results

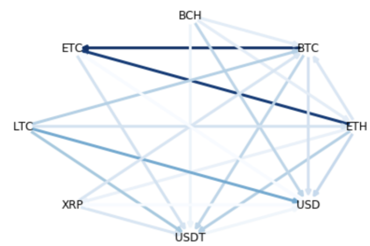
First, we simply looked at the z-scored exchange rates of our different cryptocurrencies (we only plot exchange rates with respect to BTC for graphical ease):



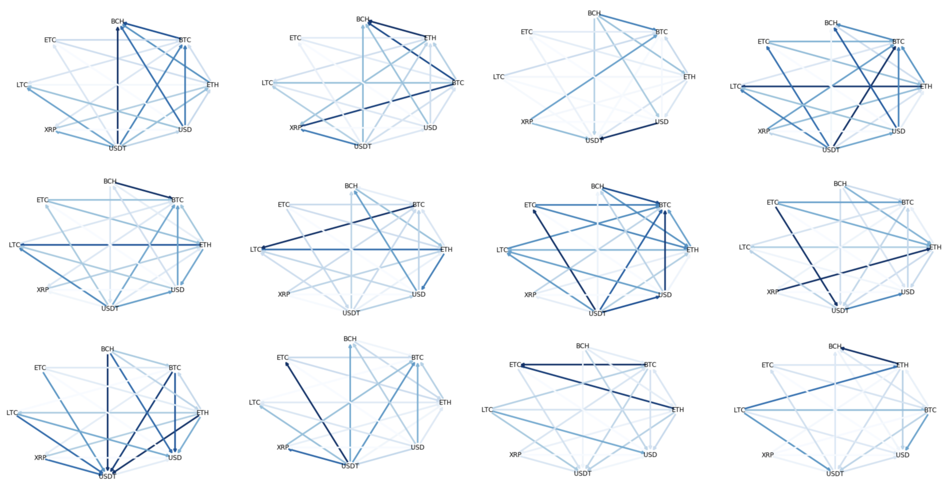
In addition, we produced correlation matrices of the ten-minute returns of our different exchange rates between June 2 and June 13 (one correlation matrix per day below). Note the sporadic behavior of these correlations as they converge and diffuse:



The first thing we noticed is that the spike in ETC/BTC on the morning of June 12 was not represented in the (relatively low) volatility of other coins. For this reason, we suspect that this is an instance of market manipulation (especially since the price is lowered right before the spike). When we consider the flow network, we see that the spike in ETC is coming almost exclusively from BTC and ETH, implying that people with BTC and ETH are pumping their money into ETC unnaturally:



Next, we plot these flows throughout the week in order to see their relationship to the exchange rates and correlation matrices:



We believe that this methodology provides an important way to view the market because one can see drivers of price changes (i.e. why relative value changes) when spikes happen like on June 12. Next, we considered the in-degree centrality and out-degree centrality of this network through time to measure buying and selling activity, respectively:

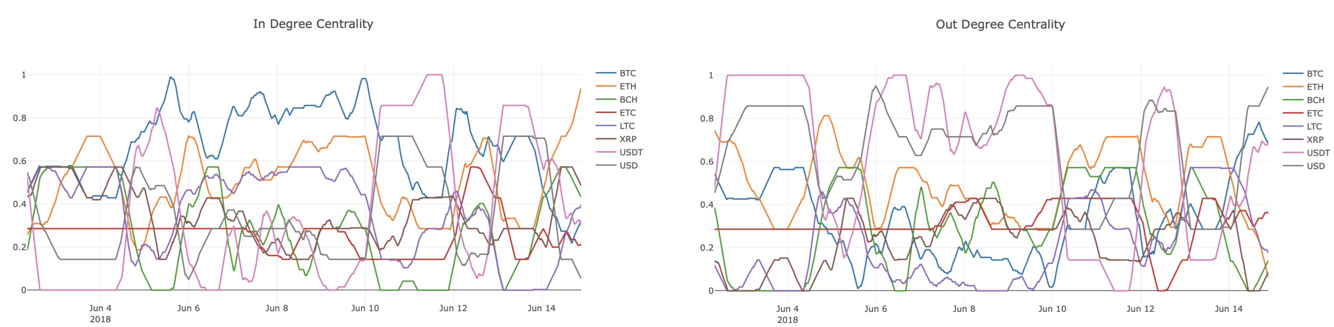
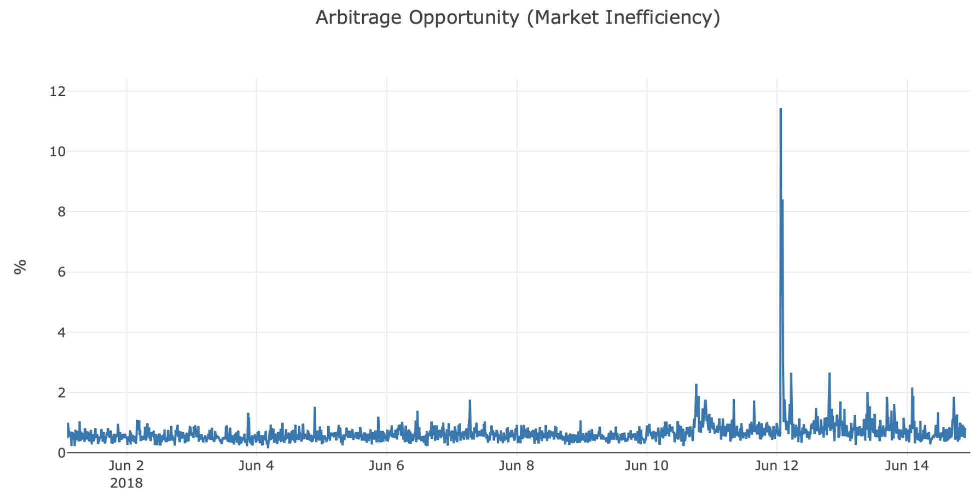


Figure 10: Eight hour moving average of in-degree and out-degree centrality of flows/momentum network.

Notice the spike in in-degree centrality (and decline of out-degree centrality) of ETC on June 12. This is further evidence that these flow graphs provide useful information.

Finally, at each point in time, we plot the maximum return of an arbitrage cycle:



The first interesting point here is that throughout time, there is a baseline level of almost 1% arbitrage opportunity, far higher than any traditional stock or FX market. Thus, we conclude that the crypto market is indeed very inefficient. The next thing to note is that, as expected, there is a very high arbitrage opportunity around the time that ETC/BTC spiked. This shows further that the price spike was not represented in exchange rates of the rest of the market, and is therefore evidence of market manipulation — it maintained a maximum return rate of 5% - 11% for roughly twenty minutes.

7 Conclusions

On an annual basis, it is evident that the devastating Bitcoin crash in January of 2018 marks a crucial pivot point in the cryptocurrency economy. More precisely, promptly after this point the cryptocurrency economy saw an undeniable correlation in the movement of prices among the major cryptocurrencies, as well as a clear centralized importance of Bitcoin.

In light of these patterns, we sought to understand how to predict the movement of currency prices. First, we identified points at which regime changes occur and hypothesize whether these are points at which profits can be made. Then, investigation by implementing a simple momentum strategy elucidates that profits are possible, and, indeed, the points of maximum profit line up with times of erratic market behavior, or regime changes. Furthermore, stable periods give more consistent but smaller positive profits. We also considered recurrent neural nets with external regressors for achieving improved predictions.

Finally, we narrow our scope to the intraday level and observe that dramatic swings occur on this level as well and are evidence of market manipulation. These times of rapid fluctuations are also moments in which the arbitrage opportunities are the greatest.

References

[1] Coin API
`https://www.coinapi.io`

[2] Coinbase API
`https://pro.coinbase.com`

[3] Alphavantage
`https://www.alphavantage.co`

[4] Coin Metrics
`https://coinmetrics.io`

[5] Cryptory API
`https://github.com/dashee87/cryptory`