Università Bocconi MILANO

20595 Business Analytics

Does the electronic world of crypto assets allow arbitrage? Can we predict revenues confidently enough to build an arbitrage-exploiting company?

Special thanks to Professor Alfonso Gambardella and Assistant Erdem Dogukan Yilmaz for lighting up our interest in scientific business analytics and helping us in research.

Team members:
Arreza Elton
Knezevic Ivan
Todorovic Nemanja

1. Introduction

Recent increase in popularity of crypto assets led us to seek profit in those markets. Crypto assets are cryptocurrencies (or cryptographic currencies). Such currencies use cryptography to secure their payment transactions. They also use blockchain technology which is a network of various different computers working together on a public ledger to record all transactions. Bitcoin, Ethereum, Litecoin, Dash, Monero, and Ripple are all examples of cryptocurrencies. There are 2071 different cryptocurrencies currently available on the market (source: Coin Market Cap; as of 22 Nov 2018). In return for their work (lending processing power for transaction verification and registration), computers are rewarded with a payment in the form of tokens. Several beliefs drove the popularity of cryptocurrencies, such as:

1) Cryptocurrencies are not subject to depreciation by inflation (unlike fiat money);
2) Cryptocurrencies are very similar to gold and they can be used as a store of value. Their value moves together with demand. If the demand for cryptocurrencies increase, the value will go up;
3) Cryptocurrencies can be used to pay for goods and services (unlike gold).

There are many digital exchanges around the world at which people can trade cryptos.

Arbitrage is a well known concept in finance which implies that the same asset has different prices at different markets. It is generally accepted that markets are efficient and that arbitrage opportunities don't exist because traders remove them very fast. However, since the crypto assets are relatively new, we want to test if arbitrage can be used as a strategy to make profits in crypto markets. We will pair the assets from diverse exchanges and check for the price differential, and then choose the pair which we want to use for analysis. Next step is to check if the price differential last long enough to be utilized, because transfering the coins from one exchange to another takes some time, and this time duration is not fixed, but it fluctuates. Therefore, our research question will focus on the following null hypothesis.

$$H_0\text{: Time (arbitrage opportunity)} <= \text{Time (transaction)}$$

We codify our transfer time as V*, and arbitrage opportunity time as v̂. Therefore, our null hypothesis can be rewritten as:

$$H_0\text{: } \hat{v} <= V^*$$

To put it in words, duration of arbitrage opportunity is lower or equal than time needed to complete the transaction, thus we cannot make profit on this trade. In case of successful rejection of the null, the alternative hypothesis is: the time the arbitrage opportunity lasts is longer than time needed to complete the transaction, and therefore we can make profit.

2. Research methodology

This research is motivated attempt to uncover profit opportunities in the crypto market. We chose crypto market because they seem to be formed in the most recent time. Other markets like the ones for equities, currencies or derivatives have been around for much longer time and continued research in those markets points out opportunities don’t exist or they are too rare.

If the opportunity is discovered, it means we can set up a company that will earn profit by repeating the same process. Process is organized in the following way:

1) Get the prices for all available assets on all exchanges;
2) Take individual asset and pair its bid and ask prices for all exchanges. Show exchange pairs where the trade could be profitable;
3) Repeat this for complete universe of assets;
4) Choose “focus asset” to proceed (in this part we narrow the analysis down to account for our limited computing power and resources);
5) Collect data on how long the opportunity lasts;
6) Collect data on how long the transfer lasts for the profitable asset pair with respect to individual exchanges;
7) Trade those asset for which opportunity lasts longer than its respective transfer time.

This simple algorithmic process would potentially generate enough cash flow, so we can focus on creating new ideas within the company which would bring even more revenue. Also, the activity of our company would increase general efficiency of the market.

We have constructed an arbitrage transaction in the following way. We use Bitcoin (BTC) to buy Ethereum (ETC) on Exchange 1 - Binance. Then we need to transfer this Ethereum to Exchange 2 - Kraken, and sell Ethereum for Bitcoin (or in other words, use Ethereum to buy Bitcoin). At the end of the transaction the ending value of our Bitcoin on Exchange 2 should be greater in quantity than our starting Bitcoin quantity on Exchange 1. This process is represented in the figure below.

Figure 1. Arbitrage transaction setup

We first thought that checking for the price differential after accounting for transaction fees will reveal opportunities, but then we realized how important is transfer time. In light of this new realization, we decided to pivot our initial idea and now check duration of arbitrage opportunities versus transfer duration. We repeat the final hypothesis once again below.

H_0: *Time (arbitrage opportunity) <= Time (transaction)* (initial writing)

H_0: $\hat{v} <= V^*$

2.1. Data collection

We have gathered data for individual outstanding offers at exchanges. Offers for each asset are characterized by their type (bid or ask), quantity and price. These offers can be summarized by market depth graph. The depth chart is a graphical representation of the pending limit orders in the order book (please see the example below).

Figure 2. Market depth for ETH/BTC

Source: Binance Exchange (as of 27 November 2018)

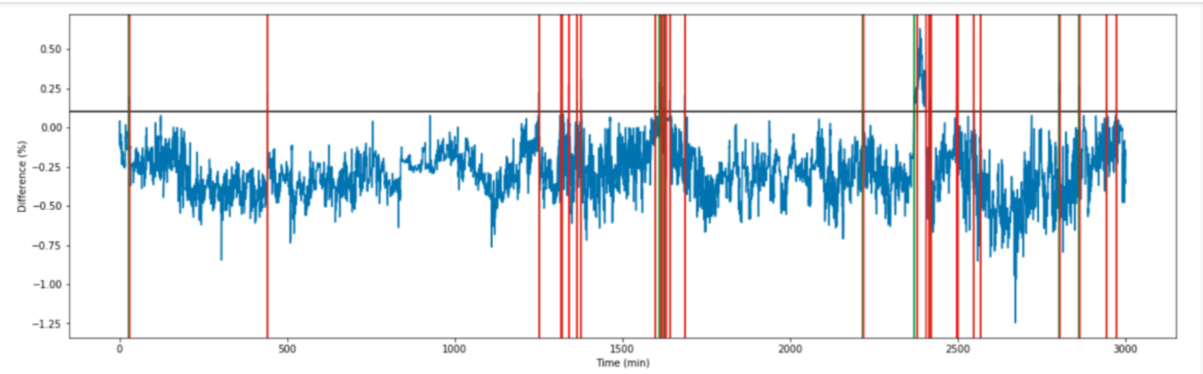
Green part of the graph represents bid offers (price and quantity at which a client can sell), while the red part shows ask offers (price and quantity at which a client can buy). Ask prices are naturally higher than bid prices.

We made a function that takes available bid and ask offers and calculates average price that would be paid for the given quantity (set by us). Then we made a function which takes as inputs quantity we want to trade, and average prices on different exchanges (pairs the asset, and then pairs the bid with ask prices on different available exchanges). Function returns price difference as a result and this data is stored in a .csv file.

2.1.1. Opportunity time: $\hat{v}$

We defined a function called frequency function. This function takes as input the data that we have collected, and returns the price differential and its time. This allows plotting the percentual opportunities on y-axis and respective durations on x-axis. As a result, we can “see” the data more clearly and think from different angles because of the visualization. We have set the threshold at 0.1% and every time the price differential would be above that, the green line would appear to indicate that opportunity has just appeared. The end of the opportunity existence is marked by red vertical line. In this way we graphically see when the opportunity appears, how long it lasts, and when does it disappear. Example of the resulting graph is below.

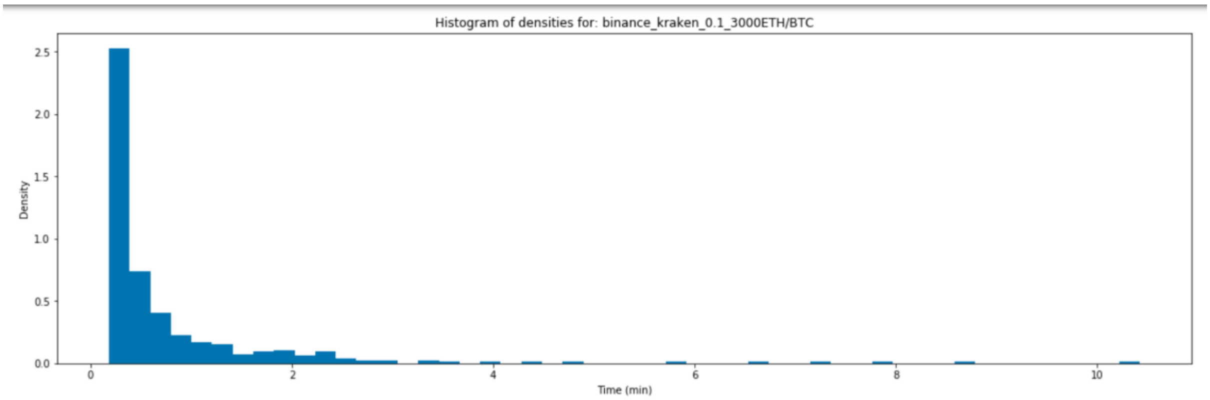
Figure 3. Opportunity time visual representation
(Obtained by running : plotpair('binance_kraken_0.1_3000', 'XMR/BTC', 0.1)



2.1.2. Opportunity time (v̂) distribution

After seeing how opportunity time looks like, next step in our analysis was to collect duration time of all opportunities that appear in a given time and to build its distribution. We have represented the distribution of opportunity duration using histogram that has time on x-axis and density on y-axis. Please see the example below.

Figure 4. Histogram of opportunity duration time
Number of opportunities: 408
Obtained by running: freq('binance_kraken_0.1_3000', 'ETH/BTC', 0)



2.1.3. Transfer time distribution: V*

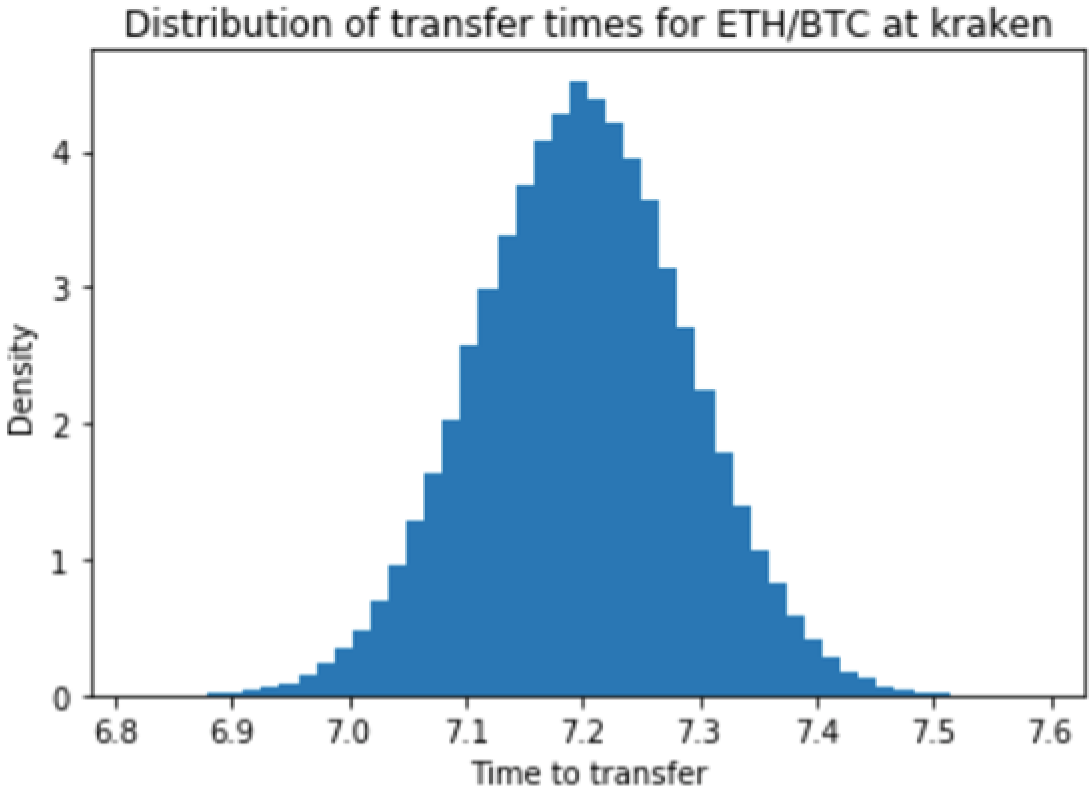
After we have collected all the information we needed about the left-hand side of our null hypothesis, we needed to obtain data representing right-hand side. Transaction transfer time is influenced by 2 main factors. First factor is the speed of the coin and its underlying technology. Without going much in-depth, some coins allow extremely fast transfer (matter of seconds, and an example could be Ripple), while some tokens take much more (for instance 10 minutes for Bitcoin). This is the time required to get confirmation for one transaction block (since the transaction is recorded in a public distributed ledger), and the block itself will be confirmed many times since many computers will verify the transfer (which improves the security of these transactions). Second factor is related to the fact that different exchanges require different number of confirmations, depending on how secure they want the transaction to be before they allow the user to see the change in the account. This differs because some exchanges want to be perceived as more secure, and by requiring large number of confirmations for a block, they decrease the chance that transactions will be reversed (which can happen in the case of fraudulent dealings). Even now it is not very easy to say what the time of the transaction will be because changes in the account happen with lower or higher confirmations number. The threshold given by the exchange is orentational. For example, the exchange may post that they require 6 confirmations for a given transaction, but in practice they can allow the change in the account after 2 confirmations, or wait until having 30. This made the analysis more complicated, but we found a way around it. We have found a database that allowed us to see mean transfer time and the standard deviation for certain exchanges, and given this piece of information, we were able to build a distribution of transfer time. Please see the example below.

Figure 5. Table of tokens and its respective mean block time, standard deviation and number of confirmations for different exchanges (Kraken, Okex, and Poloniex)

Symbol	Name	Market Cap	Mean Block Time	Std Block Time	Conf kraken	Conf okex	Conf poloniex
BTC	Bitcoin	77823843365	10.990	1.950	6	1	1
ETH	Ethereum	13908857615	0.240	0.003	30	12	30
BCH	Bitcoin Cash	4117199610	11.330	1.500	15	1	2
LTC	Litecoin	1984028729	2.860	0.240	12	1	4
XMR	Monero	1105984805	2.050	0.190	15	6	8
DASH	Dash	893429869	2.670	0.037	6	6	6
ETC	Ethereum Classic	592807463	0.235	0.004	120	30	30
ZEC	Zcash	444573706	2.500	0.059	24	15	8

Source: bitinfocharts.com (data for the past 30 days, as of 20.11.2018)

Figure 6. Distribution of transfer duration time



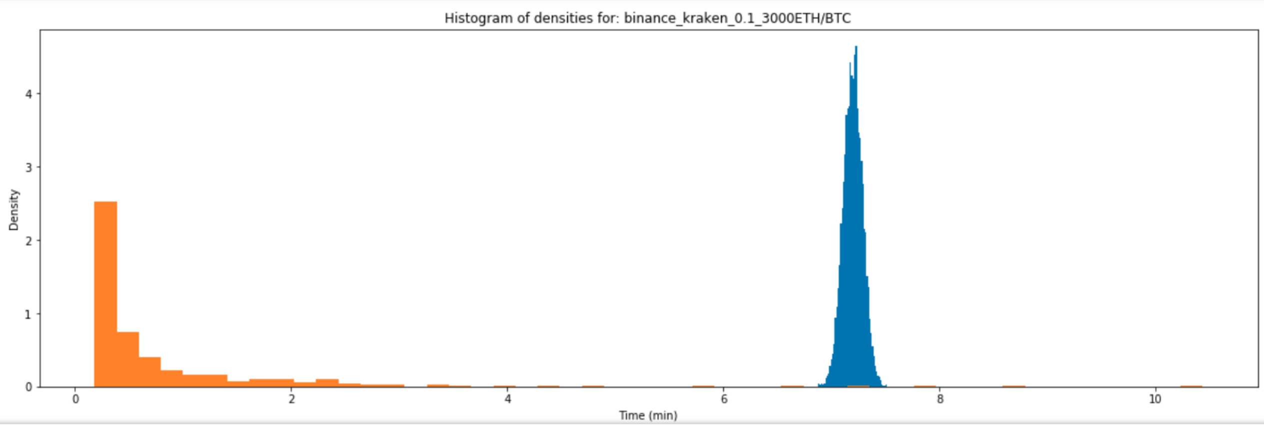
3. Results

After gathering all the necessary ingredients for testing our hypothesis, we conducted Mann-Whitney U Test on the null hypothesis. We used this specific test because unlike the t-test, it does not require the assumption of normal distributions.

In summary, we cannot reject the null hypothesis H_0: $\hat{v} <= V^*$ on any reasonable level of confidence. Please see the graph representing our findings and test below.

Figure 7. Histogram with distributions of opportunity time and transfer time on the same time scale

Obtained by running: a, b = hypothesis_tests('binance_kraken_0.1_3000', 'binance', 'kraken', freq1 = True)




```
Number of opportunities: 408
MannwhitneyuResult(statistic=4045411.0, pvalue=1.0)
```

This graph shows visually that the null hypothesis cannot be rejected because almost all available opportunities do not last long enough to transfer the coin and make profit.

4. Limitations

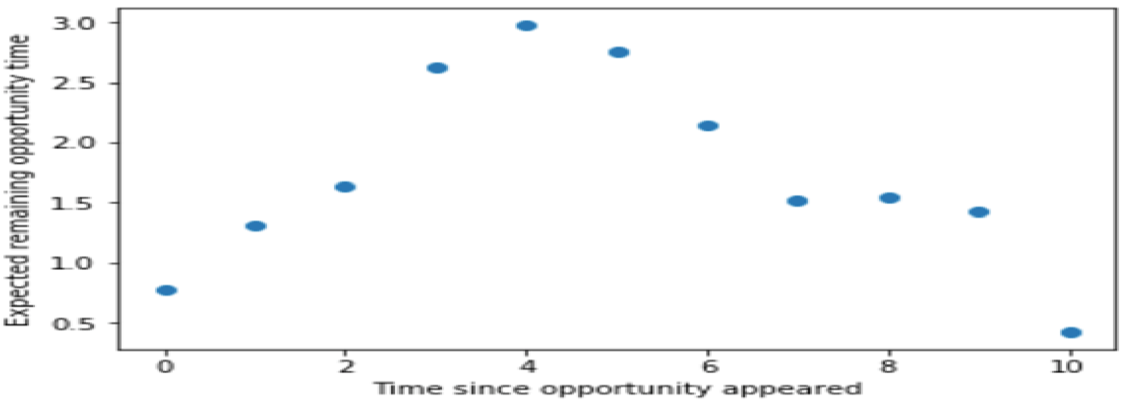
We are very excited about our research, but regardless we have to admit its limitations. Main concern that we have is a relatively small dataset that we collected and used for the analysis. Our data was collected over 50 hours continuously. We acknowledge that we need to collect data for longer period to be able to produce better and more confident results. In addition to collecting data only for a short time period, we also gathered data only for four exchanges (Binance, Okex, Poloniex, and Kraken). In order to have more complete analysis we need to incorporate larger number of markets in our model. Moreover, we have focused only on the most visible coins for which we could gather all the data needed, more specifically transfer times. We can bridge this problem by manually transferring all the coins that we want to analyze and measuring the transfer time, or by further digging in web scraping direction (from https://chainz.cryptoid.info/). Also, we believe that, for the most popular and visible coins, the arbitrage is removed fast, and because of this, our results might exhibit negative bias. Theory also suggest that arbitrage is more probable for less known coins with lower liquidity and trading volume. Another aspect that can be improved is the fact that we only checked for the simplest form of arbitrage, which includes buying at one market, transfering, and then selling at another one. While in fact there are many strategies for exploiting arbitrage opportunities, such as relative trading strategies and triangular arbitrage strategies. At the end, we did not include transaction fees in model, as we wanted to check if it is possible in the first place. If the hypothesis had been rejected, incorporating the transaction fees would be natural next step.

5. Conclusion

Even though we could not reject the null hypothesis, we take it as an opportunity to pivot and plan to use it as a starting point for further research. Since we used a small number of exchanges (only 4) to pair likewise limited number of cryptoassets (10 coins), we believe we could improve our results by using more exchanges and larger number of tradable asset, for which we will need additional resources (for computational power) and time, considering the size and experience of the team. We are considering using of Google Cloud Platform to create a virtual machine for data collection and computation.

More data could tell us more. Our meagre data suggest that gave us an idea that expected remaining opportunity time might be a function of time since the opportunity has appeared, and thus can be maximized. Please see the example below.

Figure 8. Expected remaining opportunity time as a function of time since the opportunity has appeared (brackets below the graph indicate number of remaining opportunities)



[408, 83, 33, 13, 8, 6, 5, 4, 2, 1, 1]

In case additional data supports this assumption, we could find optimal waiting time and thus have a greater chance of rejecting null hypothesis.

Very naive approach to interpret the Figure 8 is: if we waited 4 minutes to start the execution of the arbitrage transaction (instead of doing it immediately), we could expect to have 3 minutes to transfer ETH to another exchange and make profit, which current data suggest as an optimal waiting time. If exchange that we are transferring to is for example Okex (which requires 12 confirmations), instead of Kraken (which requires 30 confirmations), we could expect the distribution of transfer duration time to have mean around 3 minutes. We believe it would be very interesting to test the hypothesis for this case. However, it is crucial to get more data to test this.

6. Appendix

For all details regarding the research, please see the the code script and the data.