1 Abstract

As the cryptocurrency craze continues to spread, billions of dollars are exchanged in the world of digital currencies daily, across hundreds of exchanges. These exchanges often have significant disparities in price. Thus, the unregulated market provides the potential opportunity for interexchange arbitrage as a method of profit.

For my senior project, I seek to examine the feasibility of this arbitrage at any given point in time, looking at data from several major cryptocurrencies and exchanges. I will use machine learning that looks to train a model to calculate the expected gain in dollars after trading and transferring across exchanges. I will also train a model on foreign currency exchanges to serve as a control. As there should be relatively little market inefficiency in the foreign exchange market, the cryptocurrency arbitrage model should discover more arbitrage opportunities.

To test, the trained models will predict historical data, and the expected gain spit out by the models will be compared to the actual profit percentage. This will ascertain the feasibility of the model; once tested, the model will be able to determine the current situation of cryptocurrency arbitrage.
2 Background

The cryptocurrency market is still very new and unregulated; as a result, one of the tactics difficult to use on Wall Street—which is heavily regulated—but much easier to find opportunities in cryptocurrency is arbitrage, especially interexchange arbitrage. Due to factors such as high volume, high volatility, and slow transfers, the price of a cryptocurrency across different exchanges can be severely out of line. In November and December of 2017, the price disparity between two different exchanges would sometimes hit 5-10%, a figure inconceivable on today’s Wall Street.

Trading on each cryptocurrency and exchange constantly fluctuates as seen by volume, and as a result often so does the price disparity. For example, on occasion the price disparity of Bitcoin on the exchange GDAX compared to Bittrex will be 10%. At other time points, it will be 0.1%, while the price disparity of Ethereum on GDAX compared to Kraken will be 8%. Sometimes, this price disparity lasts a couple seconds or minutes, before the blockchain can even confirm a transfer, and other times it will last for hours. In addition, blockchain confirmation time can be erratic depending on volume, as well as the price of the cryptocurrency coin itself. If the transfer process is too long, the price can shoot up or down due to the volatility of cryptocurrencies, essentially cancelling out any potential arbitrage profit. I seek to take into account all these factors in order to train a model to determine the feasibility of arbitrage at a given point in time.

3 Outline

1. Gather cryptocurrency data. Collect ticker information that each exchange’s API provides by researching documentation. Start with the exchanges with highest volume that US residents can trade on (GDAX, Gemini, Kraken, Bittrex, etc.). Variables such as volume and market cap could potentially be key predicting features but some exchanges may not provide this information. In addition, collect confirmation time data—how long currency is stuck on the blockchain mid-transfer.

2. Find reliable historical Forex data on the same time interval as the cryptocurrency data.

3. Calculate the arbitrage profit (or loss) that would have occurred based
4. Train many long short-term memory networks (or other more suitable model, additional research necessary) on the data, given features deemed important as input and the calculated arbitrage profit as training output. Test different training networks for specific cryptocurrency pairs and exchange pairs, as well as a model including data from all cryptocurrencies and exchanges. For example, the network modeling arbitrage profit for the path USD (GDAX) → BTC (GDAX) → BTC (Bittrex) → ETH (Bittrex) → ETH (GDAX) → USD (GDAX) could potentially be very different if using different cryptocurrencies or exchanges.

5. Run regression / perform data analysis to determine the accuracy of the models, and adjust input features or data models used if necessary.

6. Train a model for foreign exchange data as well and compare with the cryptocurrency models.

7. Create a command line application that determines the highest expected arbitrage profit at the time of running the application. It will output the suggested trades on specific exchanges / cryptocurrencies and the expected profit. If time permits, wrap this in a web application.

4 Timeline

1. January to mid-February: Data mining + scraping
2. Late February: Calculate arbitrage for trade combinations
3. Mid-March to mid-April: Train + tune models to find best fit
4. Late April: Wrap into application
5. Early May: Write report
5 Methods

The data scraping / cleaning, calculations, model training, data analysis, and the tool will all be written in Python. Packages I plan on using include numpy and pandas for handling large datasets, as well as scikit-learn for machine learning tools.

6 Deliverables

In addition to the final report discussing my model selection and comparison between cryptocurrency and foreign exchange markets, I will provide the cleaned data used in training the model (with arbitrage), as well as either a command line or web application output the best trades to perform based on my models.