

# Micro Data Analysis and Forecasting of Cryptocurrency Markets

Alfred Holmes

September 2019

## 1 Introduction

Cryptocurrencies are typically traded on order-driven markets in which market participants place limit and market orders. Limit orders are either bids or asks, bids being offers to buy at a particular price and asks offers to sell. Market orders are orders executed against the best available limit order prices (buying at the lowest ask or selling at the highest bid). If a limit order is placed that overlaps an existing limit order on the other side of the market, for example there is an offer to buy 1 BTC for 2 USD and someone offers to sell 1 BTC for 1 USD then the second limit order will be considered by the exchange as a market order and the trade will be priced at the original limit order. Order-driven markets are now the standard way that stocks, futures and currencies are traded. Two benefits of studying cryptocurrencies over traditional financial instruments is that the exchanges freely give out discretized order data and that the markets are always open, except for the odd update here and there. In a market, the quote asset is the asset against which the market asset is traded. For example in a BTC against USD market, the quote asset is USD.

This study examines the available granular data from the cryptocurrency markets to develop statistical models which explain the mechanisms for different factors cause price movements on both a large and micro scale. To do this we estimate the marginal distribution of the properties of the next order to arrive at the exchange, given the current market state.

There are two types of trades that occur on cryptocurrency markets, high frequency (risk free or low risk) trades and trades to facilitate longer term (risky) investments. Typically high frequency trading strategies can be completely hedged and therefore risk free, but some traders may use statistical models or gambling to try to increase profits by not completely hedging. Examples of high frequency trades include market making and arbitrage. Market making is where an individual provides liquidity by offering to both buy and sell an asset. They place limit orders in such a way as to try and always buy and sell the same amount. Market makers make money by having a difference in the prices at which they buy and sell[4]. When a trade occurs on an exchange there is a maker and a taker, the maker being the party that provided the market liquidity by placing a limit order and the taker the party that is removing (taking) liquidity from the market. The presence of market makers gives meaning to which side of the market the maker placed the limit order on as typically the taker will be from a longer term investment. If, for example, there is a higher rate of market orders to buy than sell then the price is likely to increase even though on the exchange exactly the same amount of the asset is being bought and sold. The mechanism for this price increase is that the market makers will either raise their prices to sell due to the increased demand or increase their buy prices to increase the chance that a seller will sell to them and so they will reduce their exposure. Arbitrage is a different high frequency technique where traders (using automated systems) look for discrepancies in the pricing of assets to make risk free profits. This can be achieved by making a trade and on a different exchange the reverse of that trade or by trading in a loop on one exchange that offers different quote assets. For example, if the price of 1 BTC was 1 USD, 1 ETH was 2 USD and there was an offer to buy 1 ETH for 1 BTC then the arbitrage trader would spend 1 USD on 1 BTC, trade the 1 BTC for 1 ETH and then sell the 1 ETH for 2 USD, making 1 USD profit. It is through this mechanism that the markets on different exchanges keep the same prices and the prices on one exchange are self consistent, unlike the example given.

Each market on each exchange is a complex system with external influences, primarily other markets and news relating to the traded assets. This makes the study and forecasting of cryptocurrency market dynamics rather complicated as there are many exchanges each offering many markets, with any trade on any one of the thousands of markets perhaps influencing the whole system. It is reasonable, however, to only focus on the markets with the highest trading volume and assume that arbitrage trades act as a proxy for market movements that happen on other markets. The mechanism for this is that the arbitrage traders will remove equal liquidity from the two different markets and so for the larger (higher liquidity) market the effect is as if the original market move occurred on that market and on the smaller market the trading price does not move until there is consensus from both markets. It is still the case that there is a measurable effect due to arbitrage, but following this line of thinking, the analysis of this study applied to different markets should be invariant under the typical volume

traded. This will be examined by performing the analysis in this study on two exchanges with differing typical volume. As for inter-currency arbitrage the mechanism is quite similar, where arbitrage trades transfer the market movements between two markets of the same currency with a different quote asset. (TODO find references for this)

In all markets, the long term (days after an event) affects of news or developments are not trivial to forecast. In order to effectively model the market, how order rates are related to both recent and historic prices and price changes needs to be understood. Love and Pain (2008) show that there is a measurable effect of news on the observed order rates (flows) on foreign exchange markets, but the affects last only for a couple of minutes[7] and so, unless the news fundamentally changes the value of the quote asset, are unlikely to contribute to a general trend of the price of an asset and takes up a large portion of this study. We find that, after seasonal adjustment, a key factor in determining the order rate is how far away from moving averages the current price is. The speculative nature of cryptocurrencies holds significant weight in the affect of news in the cryptocurrency markets, meaning that any development in the cryptocurrency space is unlikely to fundamentally change the speculative value of the assets. Unlike traditional currencies, bonds or dividend paying stocks, there is no general consensus for how much each cryptocurrency should be worth and so the significance of any development in the space is very difficult to measure, and whether or not any market activity following a development is due to the movement in price or due to the development itself. In this study we do not look at macroeconomic factors and assume that random chance from fitted probability distributions is sufficient to replicate reactionary behavior.

The scripts written for this study to download, analyse and model the cryptocurrency market data can be found in a repository on *GitHub*.

## 2 Data

### 2.1 Orders

Data was acquired from the exchanges Binance and Coinbase, which respectively account for approximately 17% and 2.6% of the global spot volume at the time of writing.[6] These numbers are likely to have been different in the time frames studied, especially before 2018 because Binance only opened in late 2017.<sup>1</sup> Both exchanges offer REST APIs[1][5] which give historical trade data. Each market order (that is either a market order executed on an exchange or a limit order that overlaps with the other side of the order book and is hence reported as a market order) results in a series of trades where liquidity is removed from the exchange. For example if a large market order is placed that ends up removing many levels in the order book, then in the data we see multiple trades at the same time increasing in price. The exchanges record and publish the time, volume, price and the side of the order book that the maker is on. The in the github repository can be used to download this data although it takes some time due to rate limits. In order to analyse the orders, the raw data from the exchanges is downloaded and stored in a database and then orders are that occur at the same time and remove liquidity from the same side are grouped together. The `lib/` folder of the *GitHub* repository contains a *python* package `exchange` that is used throughout the study to handle database interactions in order to load orders.

### 2.2 Order Book Data

Typically cryptocurrency exchanges do not offer historic order book data through their APIs. Some websites do offer the purchase of orderbook data but the data has a low sample rate so does not have the required level of granularity. Binance and Coinbase do however offer Web Socket Streams of exchange data, so it is possible to record the order book through time. The Binance WSS API provides updates every 0.1 seconds [2] where as Coinbase gives real time updates[5]. At the time of writing Binance is active enough to potentially see tens of changes to the orderbook (reported as the `Update ID` in the order book WSS) per sample, with Coinbase enjoying about 10% the number of updates. An example script recording the WSSs can be found in the *Github* repository in the `Download/` folder. Decentralized exchanges in which orders are written to a blockchain, like Bitshares, clearly allow the access of historical order book data, but these have not been explored due to the low volume and the lack of high frequency trading on these exchanges, making them fundamentally different from the main high volume price driving exchanges studied[3].

---

<sup>1</sup>The website `coinmarketcap.com` can be used to track exchange volume through time, although the reported volumes may be inaccurate with some exchanges reporting fake volume. We haven't perused the tracking of the studied exchanges' volume through time

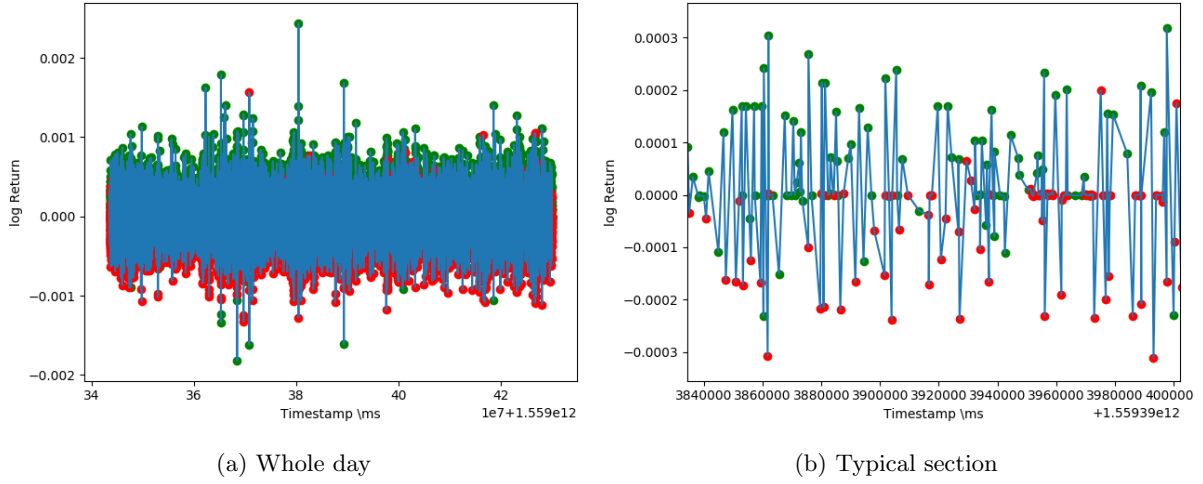


Figure 1: The log returns for each order on 1st July 2019. Green points represent market buy orders, red points represent market sell orders.

### 3 Empirical Analysis

#### 3.1 Market Movement Time Series

From each order, it is possible to calculate the price change caused by that order. That is, if the order occurs at time  $t$  and  $S_{t+(-)}$  is the last traded price just after (before) time  $t$  then the price change caused by order  $i$ ,  $R_i$ , is given by  $R_i = S_{t+}/S_{t-}$ . Let  $r_i = \log R_i$ . Figure 1 shows this data plotted for a typical trading day. We see from this that orders can generally be placed in to one of three categories. They either increase the price, decrease the price or do not alter the price. This makes a lot of sense as any incoming order order will either be on the same or opposite side of the order book, so the price will either change down, likely undoing the previous increase (leading to the observed symmetry) or it will stay the same. Sometimes an order will break through a level in the order book leading to two consecutive price changes in the same direction.

#### 3.2 Order Rates

As explained below, a key statistic when modelling an order-driven market is the order rate.

#### 3.3 Price Change Variance

### 4 Market Models

#### 4.1 Basic Markov Binomial Model

In this model, each order either increases the log price by  $\delta s$ , decreases the log price by  $\delta s$  (we ignore trades that do not alter the price). We then assume order states (whether the order increases or decreases the price) form a Markov chain  $X_i$  on the state space  $\{-1, 1\}$  with transition matrix  $P = (p_{ij})$  which is such that a stationary distribution exists. Let  $\lambda_{ij}, i, j \in \{-1, 1\}$  be the order rate for the chain moving from state  $i$  to  $j$ . By the ergodic theorem if  $V(s, n)$  is the number of visits to state  $s$  before step  $n$  and  $\pi$  is the stationary distribution and  $\sigma^2$  is the asymptotic variance, then by the central limit theorem

$$\frac{V(s, n) - n\pi_s}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (1)$$

Hence if  $U_n$  ( $D_n$ ) is the number of price increasing (decreasing) trades before time step  $n$ , we have that

$$D_n = V(-1, n) = n - V(1, n) = n - U_n \quad (2)$$

$$U_n - D_n = 2U_n - n \quad (3)$$

$$U_n \approx \mathcal{N}(n\pi_1, n\sigma^2) \quad (4)$$

for large  $n$ . We now look at the distribution of the number of trades in a large time period,  $S$ . Let  $T_l$  be the  $l$ th interarrival time (so  $T_l \sim \sum \chi_{\{X_l=j, X_{l+1}=j\}} A_{ij}^l$  where  $A_{ij}^l \sim \text{Exp}(\lambda_{ij})$ ). So in the stationary distribution,

$$\mathbb{E}(T_l) = \sum_{ij} \lambda_{ij} \pi_i p_{ij} \quad (5)$$

$$\mathbb{E}(T_l^2) = \sum_{ij} 2\lambda_{ij}^2 \pi_i p_{ij} \quad (6)$$

as  $\pi_i p_{ij}$  is the long run proportion of jumps from state  $i$  to  $j$ . Let  $\mu = \mathbb{E}(T_l)$  and  $\sigma^2 = \mathbb{E}(T_l^2) - \mathbb{E}(T_l)^2$  then if  $N$  is the number of trades from time 0 to  $S$ ,

$$\mathbb{P}(N \leq n) = \mathbb{P}\left(\sum_{l=1}^n T_l \leq S\right) = \mathbb{P}\left(\frac{\sum_{l=1}^n T_l - n\mu}{\sigma\sqrt{n}} \leq \frac{S - n\mu}{\sigma\sqrt{n}}\right) \quad (7)$$

By the central limit theorem,

$$\frac{\sum_{l=1}^n T_l - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1) \quad (8)$$

Hence,

$$\mathbb{P}(N \leq n) \approx \Phi\left(\frac{S - n\mu}{\sigma\sqrt{n}}\right) \quad (9)$$

for large  $S$ . If  $R_S$  is the log return after time  $S$ , it follows that

$$\mathbb{P}\left(R_S < \frac{\alpha}{\delta_S}\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(U_n - D_n < \frac{\alpha}{\delta_S}\right) \mathbb{P}(N = n) \quad (10)$$

which is easily approximated using the derived normal approximations.

## References

- [1] *Binance REST API*. URL: <https://github.com/binance-exchange/binance-official-api-docs/blob/master/rest-api.md>. (Accessed: 12-08-2019).
- [2] *Binance Web Socket Stream API*. URL: <https://github.com/binance-exchange/binance-official-api-docs/blob/master/web-socket-streams.md#how-to-manage-a-local-order-book-correctly>. (Accessed: 12-08-2019).
- [3] *Bitshares Exchange*. URL: <https://bitshares.org/>. (Accessed: 12-08-2019).
- [4] Tanmoy Chakraborty and Michael Kearns. “Market Making and Mean Reversion”. In: *Proceedings of the 12th ACM Conference on Electronic Commerce*. EC ’11. San Jose, California, USA: ACM, 2011, pp. 307–314. ISBN: 978-1-4503-0261-6. DOI: 10.1145/1993574.1993622. URL: <http://doi.acm.org/10.1145/1993574.1993622>.
- [5] *Coinbase Pro REST API*. <https://docs.pro.coinbase.com/>.
- [6] *FTX Volume Monitor*. <https://ftx.com/volume-monitor>. Accessed: 12-08-2019.
- [7] Ryan Love and Richard Payne. “Macroeconomic News, Order Flows, and Exchange Rates”. In: *The Journal of Financial and Quantitative Analysis* 43.2 (2008), pp. 467–488. ISSN: 00221090, 17566916. URL: <http://www.jstor.org/stable/27647357>.