Are Crypto Exchanges Wash Trading?

CS109, Jinyoung Kim

THE PROBLEM:

The world's been shocked by the fall of FTX, major cryptocurrency exchange. Due to poor accounting, internal processes around accountability, high leverage lending, and a secret backdoor connection with Alameda Research, FTX went from $32 billion company to essentially 0 in the span of a week. Even for those who don't believe crypto holds value, this has mattered. [$2 billion worth of customer funds, much of which are savings from retail investors, have simply disappeared]().

I care about this because I believe information asymmetry between retail and institutional investors is a fundamental and growing problem in our increasingly unequal world. Fun fact: my favourite movie is the Big Short – fundamentally a tale about the financial sector's Labrynthian structure and bullying of the average financial participant. Things are meant to be different in the decentralised and 'transparent' world of deFi. What if they aren't in their current design? And even worse, what if the hard-earned savings of retail investors are at jeopardy, transacted through exchanges that are fraudulent?

So, is there a way to identify fraudulent exchanges'? Specifically, we'll look for a subset of fraudulent exchanges: ones that inflate their total transaction volume (TTV) with 'wash trading', where a trader buys and sells the same asset to generate misleading market information and inflate TTV.

To do this, we'll start by thinking about what real exchanges *should* look like, and then apply those heuristics to a couple major exchanges and predict which ones wash trade. Sure, if you weren't working with FTX or Alameda closely, there was no way to predict the events of the last month. But in the future, maybe we can stay away from exchanges that are doing the transaction volume equivalent of shouting "[Bitconnect]()".

OUR HEURISTICS:

To form these heuristics, let's understand the two main types of fake exchanges. The first simply lie about displayed trades. While transactions can be publicly tracked due to the nature of the blockchain (etherscan, blockcypher etc), this can still happen. The second is to identify exchanges that practice on-chain 'wash trading'. This is when there *is* real om-chain transaction volume, but the exchange is the one buying and selling cryptocurrencies and assets to boost volume. Many of these exchanges are 'pump and dump schemes', where social media attention paired with seemingly boosted transaction volume which creates a positive feedback loop of increased real volume on the exchange before the exchange shuts down and runs away with customer funds.

Given both strategies ultimately *artificially inflate TTV*, we can form three heuristics. These are the buy and sell run lengths, trade volume, and candle size. For all three, mathematic derivations are in the appendix. We'll only consider the trading volume of $BTC since it is universal and enjoys the highest global TTV, making comparison easier. Together, these metrics should give us a good picture about the degree to which an exchange is wash trading. We will use these heuristics to assess the following exchanges. We will assume Coinbase certainly does not participate in wash trading (further justification in appendix), and compare it with the behaviour of five other top 100 exchanges (reasoning for selection in appendix).

Exchange	Coin Market Cap Ranking (29 Nov 2022)	Daily TTV ($USD) (29 Nov 2022)
Coinbase	2	1,297,709,600
Bithumb	13	197,762,982
BkEx	18	544,836,394
LBank	20	1,483,981,969
Bitforex	60	692,626,406
IndoEx	83	2,151,321,864

1. Buy and Sell Run Lengths (Geometric, Chi Squared)

The 'runs' of buy and sell orders in an exchange that does not wash trade should be *unequal* and *streaky*. Consider the following sequence "*BBBBBBSSSBBBSBBBSSBSBSSSS*" of buys and sells where "B" is Buy and "S" is Sell. The sequence of runs here would include “BBBBBB”, “SSS” and “BBB”. Intuitively, these runs should be unequal and streaky because a buy is not always instantly met with a sell. A uniform alternating pattern like "*BSBSBSBSBSBSBSBBSBSSBSBSBSBS*" should also draw suspicion.

We can think of the transaction type as a Bernoulli , following a geometric distribution P(X)~Geom(n, p). We can assume the p of the geometric approaches 0.5 the more wash trading occurs, since every buy has to be met with a sell. An organic p does not necessarily approach 0.5. Additionally, we will assume a buy and sell are I.I.D (which is obviously false – elaborated on in appendix). We will compute p (or the prior that a transaction is buy) by taking the mode of the Beta. We will then run a chi squared test to measure how close the results are to normal run lengths.

2. Trade Volume (Exponential MLE)

The size of trades in an exchange that does not wash trade should follow a *continuous, decaying distribution*, and should *mostly be rounded*.

We can assume a continuous, decaying distribution since most trades happen in low volumes, and trading volume should follow a continuous distribution. We can imagine fake exchanges might not have low transaction volumes because they want to maximise TTV. To compute the distribution of trade volume, we will fit a distribution and then use MLE to fit the lambda parameter to an exponential distribution, then detect abnormal behaviour that could signal wash trading. An exponential distribution is appropriate as most trading volume should be low amounts.

We can also assume a significant portion of trade sizes should be rounded values by assuming many real trades are conducted by real people who like to trade at ‘clean’ rounded quantities. By rounded, we mean people will trade 0.8BTC, not 0.821272 BTC. There should be at least mild spikes in volume at 1.00 BTC, 0.6 BTC, and 0.1BTC. We will arbitrarily consider 2 decimal places to be a clean rounded quantity.

3. Candle Size (Normal MLE)

A “candle” in an exchange is the total volume transacted (positive and negative) in a 5 minute interval. We can assume an exchange without wash trading should exhibit a normal candle distribution. In a market with no general sentiment, this distribution should be centred at 0, and in a bear market, it should be slightly negative and vice versa. Since the data being used is taken in the span of mid November 2022, we anticipate a negative centre for our normal, then plan to normalise it. This normalised distribution should be symmetric and be similar to other exchanges. That is, we will fit a normal distribution and use MLE to optimise parameters for Coinbase, and use that to determine abnormal behaviour from other exchanges indicative of wash trading. Again, using the derivation in the appendix, we can compute the similarity of the mean and variance with Coinbase and then compute their KL divergence as a measure of similarity.

RESULTS:

It is likely BkEx, IndoEx, and LBank are participating in wash trading. Given the crude nature of this analysis, it is hard to make further claims. Here is an overall summary:

🟢: Normal, 🟡: Abnormal/Suspicious, 🔴: Highly Suspicious

	Intuition 1: Chi-Squared Test Value	**Intuition 2.1: Trade Volume Parameter “distance” to Coinbase**	**Intuition 2.2: Total # of Rounded Trade Volume**	**Intuition 3: Candle Volume Parameters KL Divergence**	**Cumulatively: Does this exchange wash trade?**
Coinbase	🟢 6.2310	🟢 0	🟢 782	🟢 0	❌
Bitforex	🟢 6.2310	🟡 0.767	🟡 68	🟢 0.0002	❌
Bithumb	🟡 20.5671	🟢 0.233	🟢 340	🟡 0.0735	❌
BkEx	🔴 70.0629	🔴 2.767	🟡 215	🟡 0.0990	✅
IndoEx	🟡 19.02310	🔴 9543.024	🔴 0	🟡 0.1296	✅
LBank	🔴 46.0596	🟢 0.233	🔴 28	🔴 3.818	✅

1. Heuristic 1: Buy and Sell Lengths

Exchange	x^2 value	p value
Coinbase	6.2310	4.3748e-1
Bitforex	7.4762	1.3017e-4
BkEx	20.5671	0
Bithumb	70.0629	1.3658e-3
IndoEx	19.02310	0
LBank	46.05958	0

Firstly, the p values of BkEx, IndoEx, and LBank are essentially 0, meaning the null hypothesis does not hold. Secondly, these were the three exchanges that went over 20 in the chi^2 value, meaning there could be deviation of normality. From these, there should be large skepticism toward LBank, and some skepticism for IndoEx, BkEx and Bithumb.

As expected, real exchanges approximated the Geometric PMF more closely than exchanges that might be wash trading. This is because wash trading disproportionately increases the number of short interval trades due to the risk of slippage. Graphs in the appendix (Fig 1.0 – 1.5) compare the Geometric PMF with the actual run lengths of each exchange.

2. Heuristic 2: Trade Volume

Exchange	Fitted Values (λ * 1000)	abs(Distance)
Coinbase	1.233	0
Bitforex	2.0	0.767
BkEx	1.0	0.233
Bithumb	4.0	2.767
IndoEx	9544.257	9543.024
LBank	1.0	0.233

Quantitatively, IndoEx raises major suspicion, and every other exchange does not raise much suspicion. However, qualitative examination of the graphs reveal other suspicious behaviour. IndoEx does not map to the exponential distribution at all, with a trivially absurd lambda. All of its volume is also concentrated between 3 and 6 BTC which is intuitively unrealistic (Fig 2.0.4). Fig 2.0.1 shows Bitforex raises minor suspicion due to the tail being very long, with a few suspiciously huge transactions (order of 1000BTC +). However, its lambda remains low.

Fig 2.0.2 shows Bithumb follows the exponential distribution almost identically. BkEx (Fig 2.0.3) raises suspicion, with significant deviation in lambda and an unusually large number of transactions of 1.5 – 3 BTC. It also breaks continuity.

For 'rounded values' (values that cleanly end in 2dp or less) we can create a bar plot for each exchange and compare. As confirms with our intuition in Fig 2.1.0, Coinbase had a large number of 0.1BTC transactions (roughly 1% of their total count). In total, 782 transactions, or 15.6% of all transactions, were rounded. Most other exchanges demonstrate suspicious behaviour. According to Fig 2.1.1, Bitforex only has 68 transactions out of 5000, or 1.36%, that were rounded - less than a tenth of Coinbase's proportion. While according to Fig 2.1.3 there were 215 rounded transactions on BkEx, the distribution of them is clearly suspect. They are uniformly concentrated in 0.01 and 0.02, suggesting deliberate manipulation. According to Fig 2.1.4 there are simply 0 rounded transactions on IndoEx - potentially a sign of 0 real customer activity. Finally, according to Fig 2.1.5, LBank only has a tiny 28 transactions that are rounded, or 0.6% of their 5000 transactions. This too, is suspicious.

Overall, from the trade volume, IndoEx and BkEx raise high suspicion, and Bitforex, LBank raise suspicion. From the rounded value count, IndoEx raise high suspicion, and LBank, BkEx, and Bitforex raise suspicion.

3. Heuristic 3: Candle Size

Exchanges	Fitted μ	Fitted σ^2	KL Divergence to Coinbase: $KL(p,q) = log\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2+(\mu_1-\mu_2)^2}{2\sigma_2^2} - 0.5$
Coinbase	0.03998187	1.36001449	0
Bitforex	0.04328145,	1.36098091	2.0505 * 10 ^-4
IndoEx	0.07868139	1.68467228	0.0734945358
Bithumb	0.12040285	0.558623	-0.09903597932
BkEx	-0.02118488	0.25976902	-0.1295961909
LBank	-0.11289942	13.05990005	3.818463159

The KL divergence score can generally be interpreted between 0 and 1. If it reaches or exceeds one, it means the distributions behave so differently the expectation of the second distribution given the first distribution (Coinbase) approaches 0. This means LBank demonstrates absurdly unrealistic behaviour. IndoEx, Bithumb, BkEx also all warrant suspicion. Both Bithumb and BkEx, on the other hand, have unrealistically narrow spreads, lending them to suspicion (between 19% ~ 41% of Coinbase's). This could mean many buys and sells are cancelling out in a manner that is artificially identical.

CONCLUSION:

Overall, there is strong evidence of behaviour suggestive of wash trading in the major exchanges BkEk, IndoEx, and LBank. If this is true, this is highly concerning. While it's not possible with current analysis to gauge the amount of real customer assets are being traded and kept on the exchange, it is likely in the hundreds of millions of USD considering they are still major cryptocurrency exchanges that exhibit ordinary behaviour (i.e. real trading behaviour) to some extent. Since this is a simplistic analysis, I've listed several potential improvements in the appendix below. Hopefully further, more rigorous investigation is done in the area.

APPENDIX:

1. **Method and Weaknesses:**

For full code and repo, check out:
💡 https://github.com/jinyoungkim927/Wash_Trading_Exchange_Analysis/tree/main
Here are additional comments and details on the method:

a) How exchanges to investigate was determined
To assess wash trading participation, we need an exchange that confidently does not participate in wash trading. For the purposes, I have chosen Coinbase, since it has been audited very thoroughly, including a public subpoena, and demonstrating proof of reserves. This is a core assumption that lies at the basis of this report. I chose to investigate the following exchanges from a combination of high global ranking score (which Coinmarketcap computes by assigning weights to web traffic, average liquidity, volume, and the confidence of reported volume) , high DTV, and because orderbook data and trading history were accessible through making calls to CoinAPI.*

b) Choices regarding the size of data
Ideally, bigger chunks of data should be analysed, and across multiple time horizons, to have stronger faith in the conclusion. In terms of 'bigger chunks' of data, I am referring to specifically two things. Firstly, the string length of the trading history for each exchange where "B" denotes buy and "S" denotes sell was limited to 2000. An ideal investigation would use longer strings, and multiple strings, but this was not possible due to the free CoinAPI plan being rate limited. Secondly, analyzing candles should take in 5 minute intervals, but again the API rate limit, coupled with my personal computer being unable to perform computations on it, prevents this from happening - hence the choice of candles artificially being 10 seconds.

c) IID assumption in heuristic #1 is highly incorrect
I.I.D was assumed so the maths is cleaner. However, this is fundamentally untrue and potentially helps explain a discrepancy in the methodology. Computing raw p using the Beta yields a p value that is too large - plotting the distribution reveals it overshoots the true data. Let p be the value yielded by finding the mode of the Beta. Let p' be the value that optimally fits the data (discovered using MLE). Surprisingly, p'/p is not 1, and is in fact closer to 0.7. This is likely because p(Buy) = f(price), p(Sell) = f(price) , so the underlying demand and supply of BTC is not being factored in when considering this - the data was taken during a time generally considered a 'bear market' where more sells were occurring at a higher volume. Constructing a demand & supply model could help with this.

d) Conclusion-consequential data cleaning decisions
In reality, calling historical trading history produces a JSON file where the trade type is "BUY", "SELL", "POTENTIALLY_BUY", "POTENTIALLY_SELL", "UNCERTAIN". I have squashed the potential buys as buys, which assumes high conviction, and likewise for sell.

2. **Local Areas of Further Investigation:**

I could have done spread analysis. From an initial exploration of the data, I suspect spreads on suspicious exchanges will demonstrate anomalous patterns, such as central tendencies that hover around an unusual fixed amount, or spreads that stay fixed for extended periods. As opposed to this, spreads on well-known exchanges should show a consistent pattern of anchoring on zero with random variability, with periodic spikes to reflect momentary surge in volatility and change of the order book.

Given the prior of an unusually large trade, I could have also found the likelihood that a similarly large trade immediately matches it in the next couple transactions could be investigated.

3. **Non-Local Areas of Further Investigation**

a) Inference on specific amount of wash trading:
Through this information, we've only arrived at a good idea of whether wash trading might, or might not happen by spotting egregiously abnormal behaviour. However, we don't have a prediction for the volume figure that has been wash traded.

b) Non-CoinAPI accessible exchanges
CoinAPI manually scrapes information from exchanges. However, there were many exchanges that exhibited suspicious behaviour on first glance I couldn't get data for. The prime culprits include BaseFEX, ZBG, BTC-Alpha, Coinhub, Valr, and Probit.

c) DeXs
Cryptocurrency trading volume could dramatically move to DeXs in the future given the recent lack of faith in CeXs and scaleability improvements in DeXs. No DeXs were examined in this analysis. The market share of DEXs has slowly been creeping up to that of spot–there are even days where Uniswap, the largest DEX, has more trading volume than Coinbase.

d) Perpetual Futures
It would be interesting to look at the potential for perpetual futures to be manipulated. Given the novel nature of the product (only created in 2016), this would be interesting.

4. **Mathematical Derivations:**

a) Buy and Sell Run Lengths

Since the PDF of a Beta is given by:

$$f(x) = \frac{(x-a)^{p-1}(b-x)^{q-1}}{B(p,q)(b-a)^{p+q-1}}$$

The mode is given by:

$$\frac{p-1}{p+q-2}$$

This gives us the value of p.

To compute the chi squared test, we use the formula:

$$x^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Finally, we can flatten the cases of where the expected value is less than one by bucketing it, since that cause the chi squared test to error.

b) Trade Volume MLE

Since the likelihood of a PDF of an exponential distribution is:

$$\mathcal{L}(\lambda, x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i, \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

We can solve for lambda by considering

$$\frac{d \ln(\mathcal{L}(\lambda, x_1, \ldots, x_n))}{d\lambda} \stackrel{!}{=} 0$$

Giving us:

$$\frac{d \ln(\mathcal{L}(\lambda, x_1, \ldots, x_n))}{d\lambda} = \frac{d \ln\left(\lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}\right)}{d\lambda}$$
$$= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

And therefore the lambda value:

$$\lambda = \frac{n}{\sum\limits_{i=1}^{n} x_i}$$

c) Candle size (Normal MLE)

We can derive the MLE of the Gaussian by starting with the definition of argmax, where we are adjusting θ, the set of parameters from function $f(\boldsymbol{X}, \theta)$ where $\boldsymbol{X}$ is some data.

$$\theta_{MLE} = \underset{\theta}{\operatorname{argmax}} p(\mathbf{X}|\theta)$$

After splitting into each parameter (computing argmax with respect to both μ and $\sum$, using log likelihood, we arrive at:

$$\mathcal{LL} = \frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2$$

We then need to compute μ and σ^2 , which (after simplification), gives us:

$$\mu = \frac{1}{N}\sum_{n=1}^{N} x_n \qquad \sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2$$

Finally, we compute their KL divergence by using:

$$KL(p, q) = \log\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$$

5. **Figures:**

a) **Run Lengths (Fig 1.0 – Fig 1.5)**

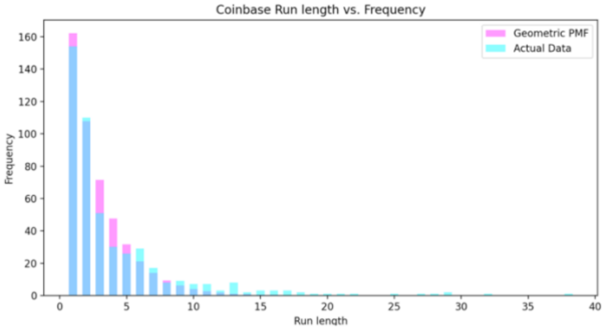

Fig 1.0

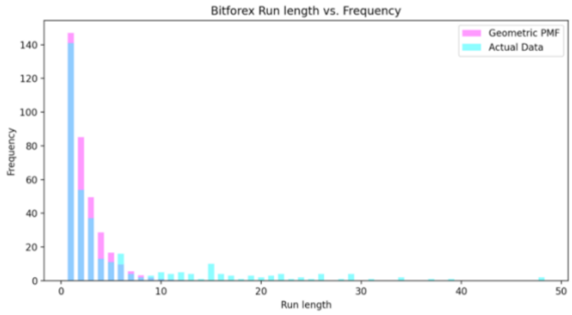

Fig 1.1

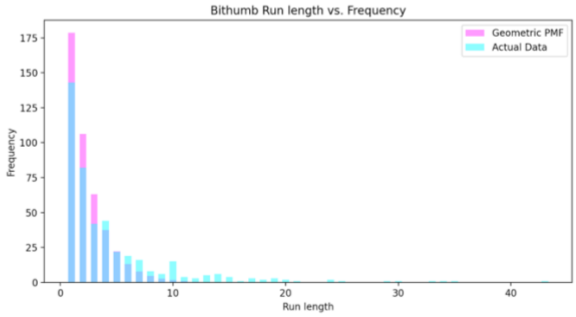

Fig 1.2

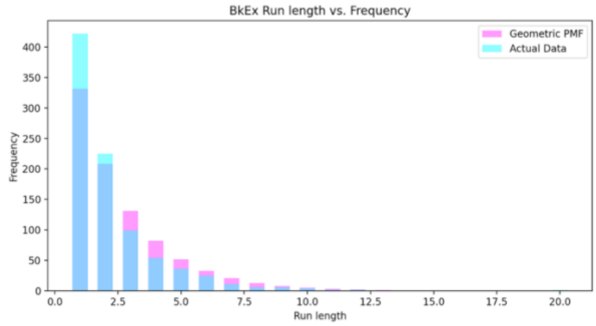

Fig 1.3

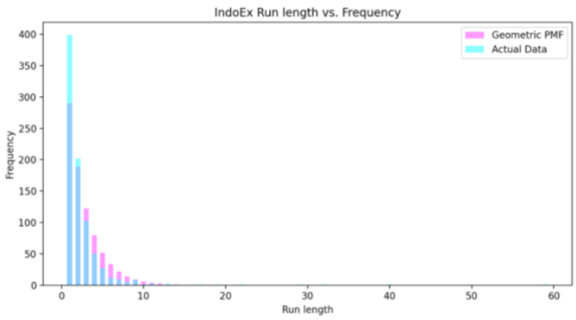

Fig 1.4

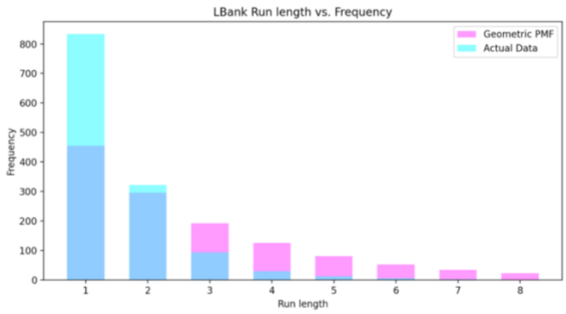

Fig 1.5

b) **Trade Volume**

i) (Fig 2.0.0 – Fig 2.0.5)

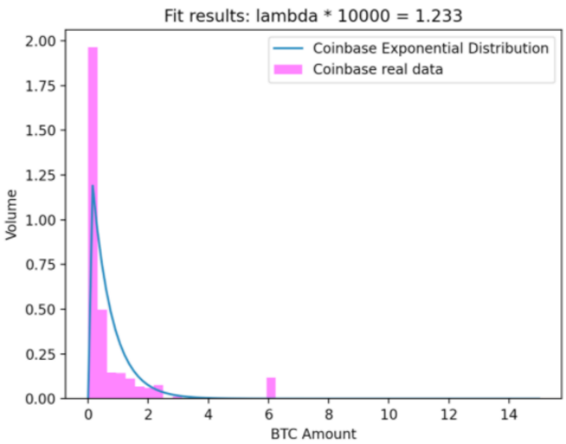

Fig 2.0.0

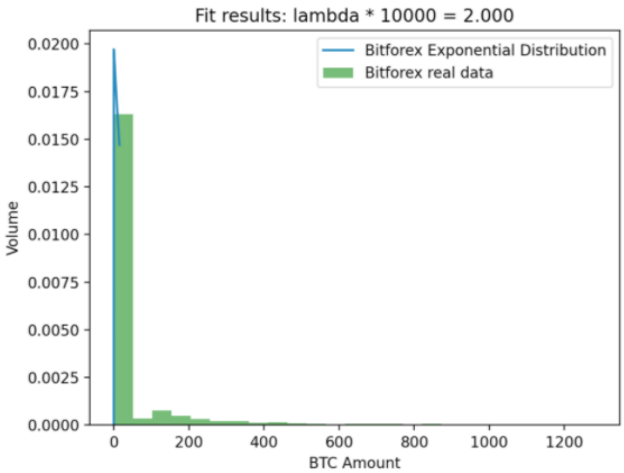

Fig 2.0.1

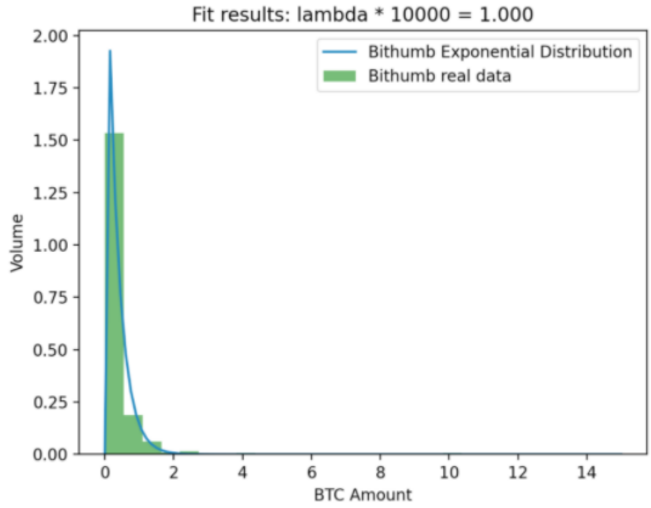

Fig 2.0.2

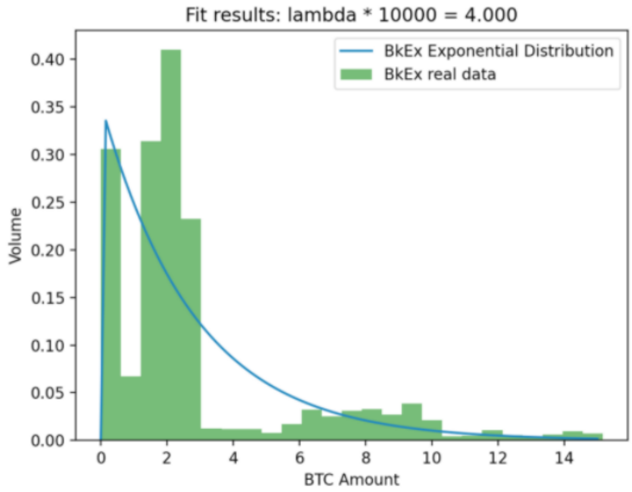

Fig 2.0.3

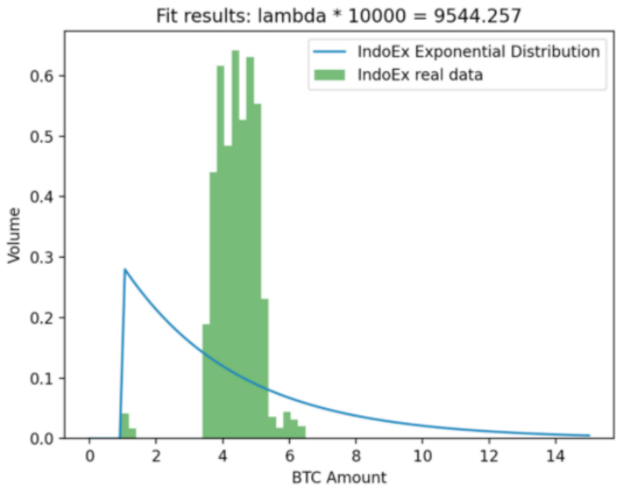

Fig 2.0.4

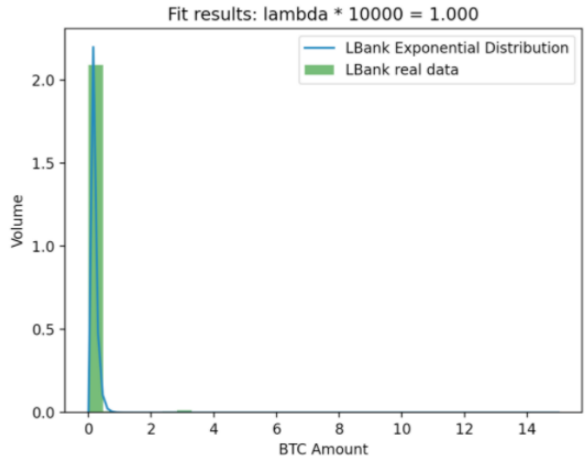

Fig 2.0.5

ii) (Fig 2.1.0 – Fig 2.1.5)

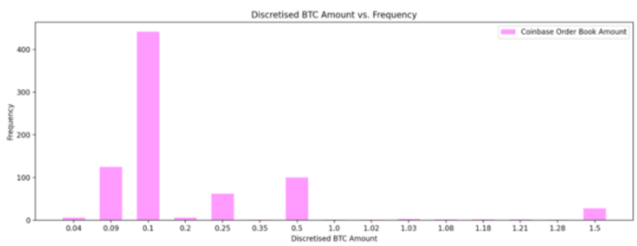

Fig 2.1.0 Coinbase

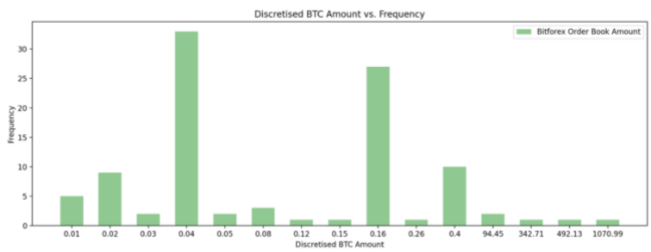

Fig 2.1.1 Bitforex

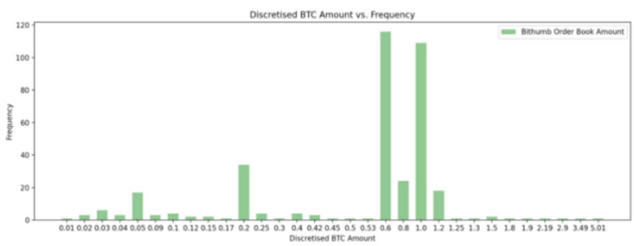

Fig 2.1.2 Bithumb

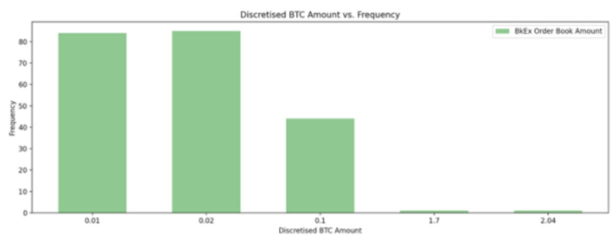

Fig 2.1.3 BkEx

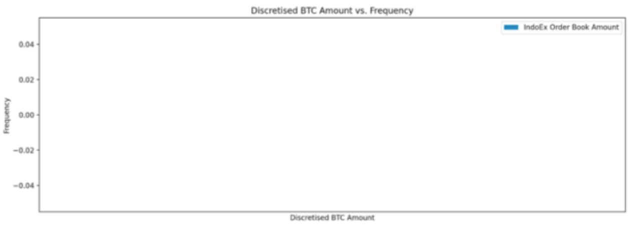

Fig 2.1.4 IndoEx

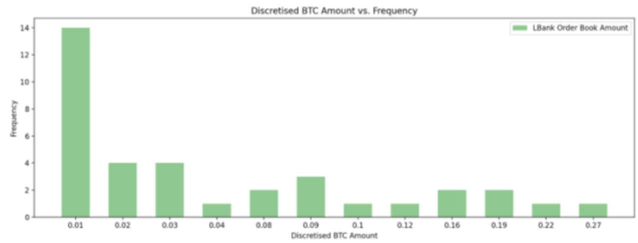

Fig 2.1.5 LBank

c) **Candle Size (Fig 3.0 – Fig 3.5)**

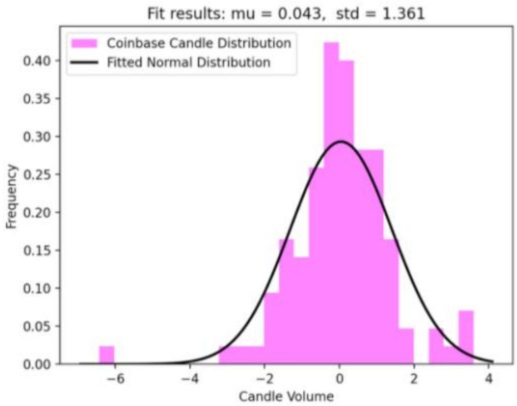

Fig 3.0 Coinbase

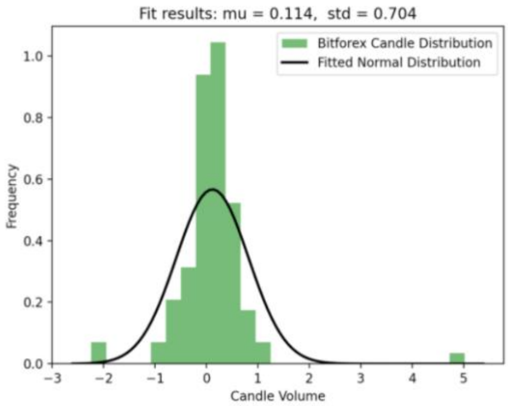

Fig 3.1 Bitforex

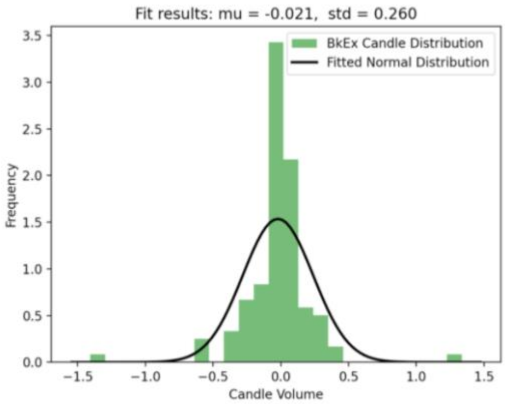

Fig 3.2 BkEx

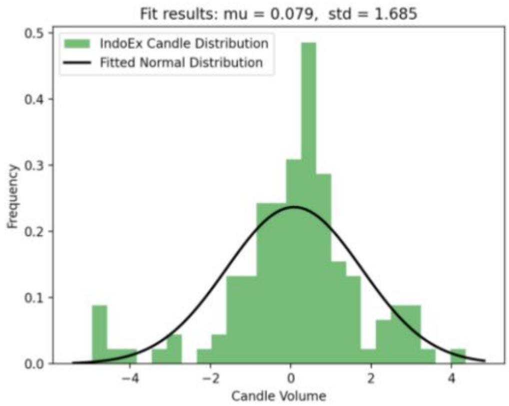

Fig 3.3 IndoEx

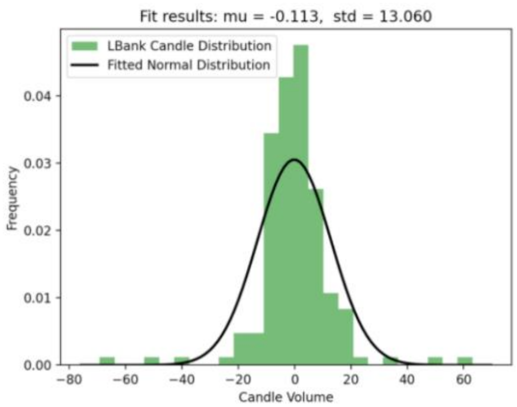

Fig 3.4 LBank

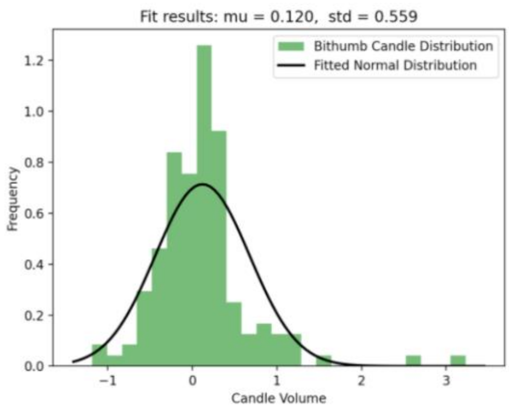

Fig 3.5 Bithumb