Wash Trading in Crypto Exchanges

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.

If you're a crypto skeptic, you should care about making exchanges more transparent so that unsophisticated retail investors aren't scammed out of their life savings. If you're a crypto believer, you should also care so decentralisation is achieved meaningfully: through disrupting 'rent-seeking industries' instead of being scammy.

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

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. Admittedly, this is rare since transactions can be publicly tracked due to the nature of the blockchain (etherscan, blockcypher etc), but it still happens. The second is to identify exchanges that practice 'wash trading' - which is more common. This is when there *is* real transaction volume, but the exchange is the one buying and selling cryptocurrencies and assets on itself 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 mix of buy and sell, trade volume, and candle size - we'll elaborate on these individually. We'll only consider the trading volume of $BTC since it is universal and enjoys the highest TTV globally, making comparison easier across exchanges. Together, they 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, with Coinbase certainly not participating in wash trading (please refer to appendix) - we will investigate several top #1000 exchanges. Further reasoning for the choice of exchanges is in the appendix. I have chosen to not use Binance since they have not been audited in the past.

Exchange	**Coin Market Cap Ranking (29 Nov 2022)**	**Daily TTV ($) (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. **Mix of buy and sell (Geom test)**

The ‘runs’ of buy and sell orders in an exchange that does not wash trade should be *unequal* and *streaky*. Consider the following sequence of buys and sells where “B” is Buy and “S” is Sell:

```
"BBBBBBSSSBBBSBBBSSBSBSSSS"
```

The sequence of runs here would be “BBBBBB”, “SSS”, “BBB” and so on. Intuitively, these runs should be unequal because each buy is not always instantly met with a sell, making the runs both unequal and streaky. A constant alternating pattern like the following should draw suspicion:

```
"BSBSBSBSBSBSBBSBSSBSBSBSBS"
```

We can think of the transaction type as a bernoulli , following a geometric distribution $P(X) \sim Geom(n,p)$. We can assume that the p of the geometric approaches 0.5 the more wash trading occurs, since every buy has to be met with a sell. This is not necessarily true generally speaking. Additionally, while we cannot assume that buys and sells are I.I.D since a buy decreases the likelihood of a following buy, we will proceed with that assumption for now (this is an important caveat discussed in the appendix). We will compute the best p (or the prior that a transaction is buy) by taking the mode of the Beta, where here it is a distribution over the probability of a Bernoulli random variable. Since the PDF of a Beta is given by:

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

and the mode is given by:

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

we can compute this from the parsed string.

We can then compare the degree to which different exchanges exhibit normal sequences of buys and sells by comparing the distribution of run lengths in a sequence with a chi squared test, with

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

where

$O_i =$ number of observations of type i

$N =$ total number of observations

$E_i = Np_i =$ expected theoretical count of type i

$n =$ number of cells in the table

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. We can also get the p value from this as well.

2. **Trade Volume (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 parameters to the Coinbase, and compare the fit to detect abnormal behaviour that could signal wash trading. Here, we'll fit to an exponential distribution. This is appropriate as most of the trading volume should be low amounts with high volume. To solve for the MLE, we do the following. The likelihood function of the PDF of an exponential distribution is:

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

To calculate the MLE, we need to solve for λ in:

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

That gives us:

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

Hence,

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

We can also assume trade sizes should mostly be rounded values by assuming real trades are mostly conducted by people who like to trade at ‘clean’ rounded quantities. By rounded, we mean people will trade 0.8BTC, not 0.821 BTC. This means we should expect spikes in volume at 1.00 BTC, 0.6 BTC, and 0.1BTC would make sense. We can record the number of transactions with only 2dp or less for this.

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 centered 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.

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.

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

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:

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

Results:

It is highly likely BkEx, IndoEx, and LBank are participating in wash trading. It is also possible Bitforex is participating in wash trading. Let’s observe the performance of the exchanges. Here is a high level summary:

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

	Intuition 1: Chi-Squared Test Result: x^2 value	**Intuition 2.1: Trade Volume Parameter “distance” to Coinbase**	**Intuition 2.2: Total # of Trade Volume ending in 2dp from 5000**	**Intuition 3: Candle Volume Parameters “distance” to Coinbase**
Coinbase	🟢 623.098	🟢 0	🟢 782	🟢 0
Bitforex	🟢 747.618	🟡 0.767	🟡 68	🟢 0.003

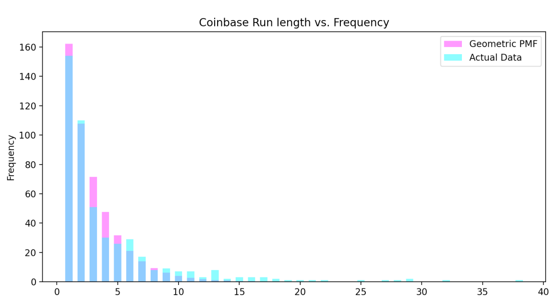
Bithumb	🟡 2056.713	🟢 0.233	🟢 340	🟢 0.3270
BkEx	🟢 700.629	🛑 2.767	🟡 215	🟡 0.805
IndoEx	🟡 1902.310	🛑 9543.024	🛑 0	🟡 1.102
LBank	🛑 4605.958	🟢 0.233	🛑 28	🛑 11.701

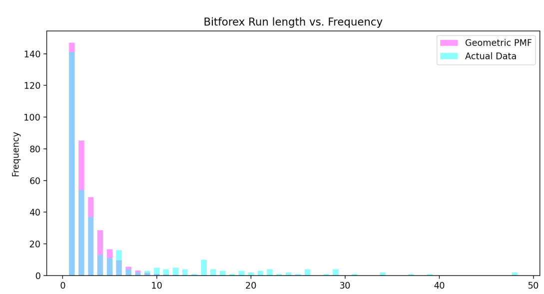
Heuristic 1: Mix of Buy and Sell

Exchange	**x^2 value**	**p value**
Coinbase	623.0977411764705	4.3748045128677945e-2
Bitforex	747.6179856115107	1.3016824359801156e-28
BkEx	2056.7134624145783	0.0
Bithumb	700.6285638297873	1.365821019259874e-18
IndoEx	1902.3096530359353	0
LBank	4605.957963105304	0.0

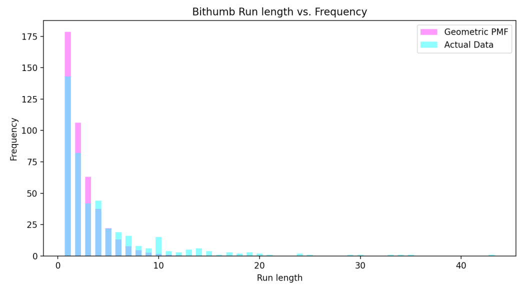
3 observations. Firstly, the p values of the findings are very low, meaning the null hypothesis does not hold. Secondly, there were only three exchanges that went over 1000 in the chi^2 value, meaning there is significant deviation of normality, and that the runs are probably not produced by random chance. From these, there should be large skepticism toward LBank, and some skepticism for IndoEx and Bithumb.

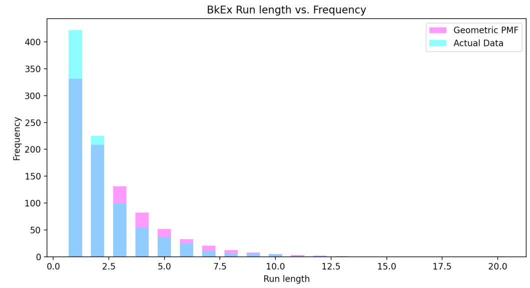
As expected, real exchanges tended to approximate the Geometric PMF much more closely than exchanges that might be wash trading. As an aside, the long tail is difficult to fit, however



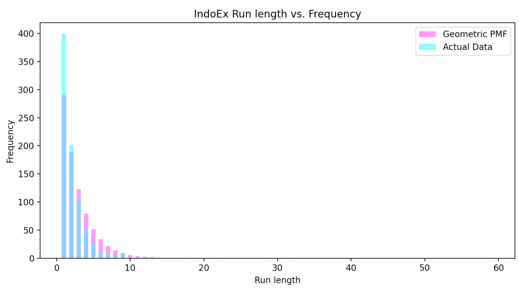


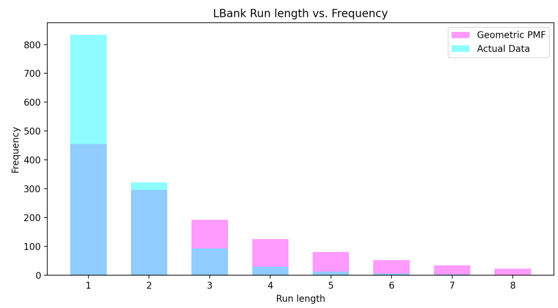
This is because wash trading disproportionately increases the numer of short interval trades, as holding onto any asset for a long period of time can create slippage and cost the holder money.





We see this suspicious behaviour mostly with BkEX, IndoEx, and most prominently with LBank.

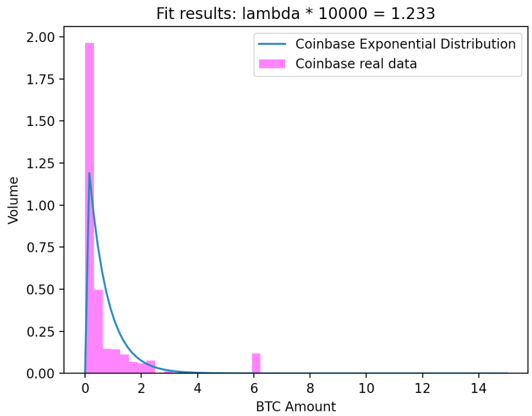


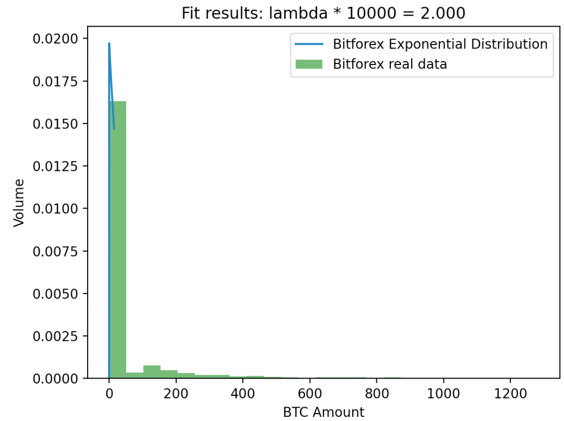


Heuristic 2: Trade Volume

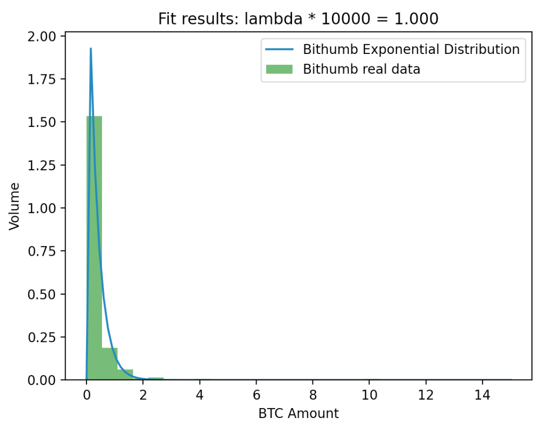
Exchange	**Fitted Values (λ * 1000)**	**abs(Distance)**
Coinbase	1.233	0
Bitforex	2.0	0.767
BkEx	1.0	0.233
Bithumb	4.0	2.767
IndoEx	9544.257	9543.024
LBank	1.0	0.233

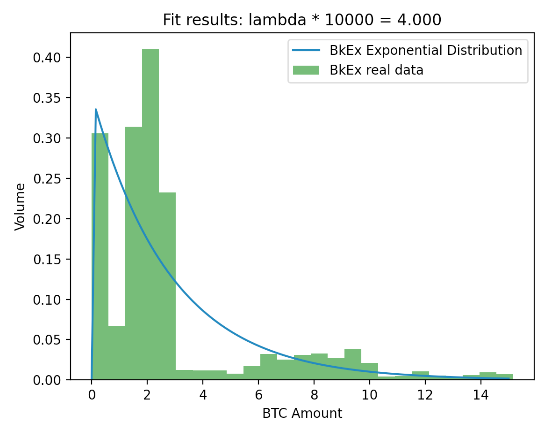
Having computed the λ parameter (multiplying by 1000 because the dataset is so horizontally sparse), we can now interpret their significance.



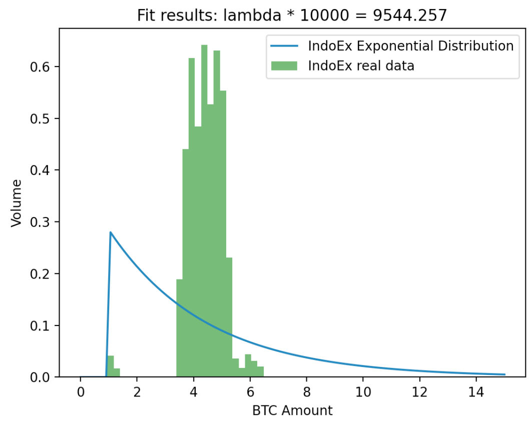


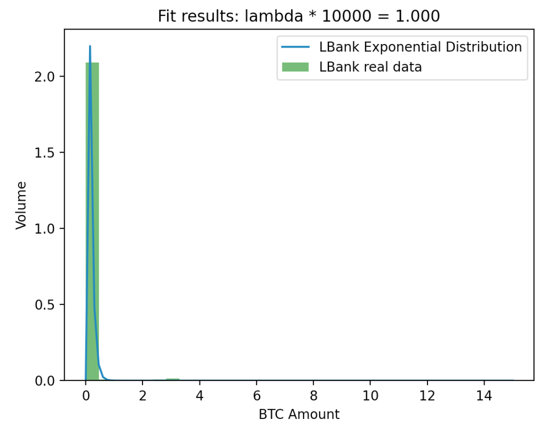
Bitforex raises minor suspicion here, due to the tail being very long, with a few huge transactions (in order of 1000BTC+) probably for the purpose of wash trading. However, its λ remains relatively low.





Bithumb follows the distribution almost identically. BkEx however raises suspicion, with a significant deviation in λ, and a pattern that features an unusually heavy load of transactions in the 1.5 - 3 BTC territory. It also breaks the continuity that we set up as a criteria.

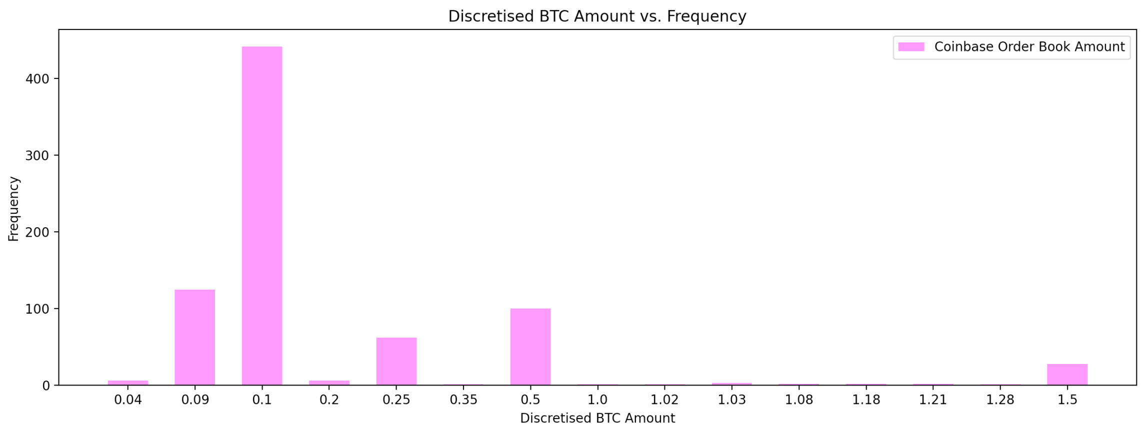




IndoEx does not map to the exponential distribution at all, yielding an absurdly trivial λ value, and being solely concentrated between 3 and 6 BTC. This is highly unrealistic since most users do not transact in the order of around $60000 USD (as of 29th November 2022). LBank also draws suspicion for being so tightly concentrated in the low bands with near zero volume with higher BTC amounts.

From the trade volume, BkEx, Indoex raise high suspicion, and Bitforex, LBank raise suspicion.

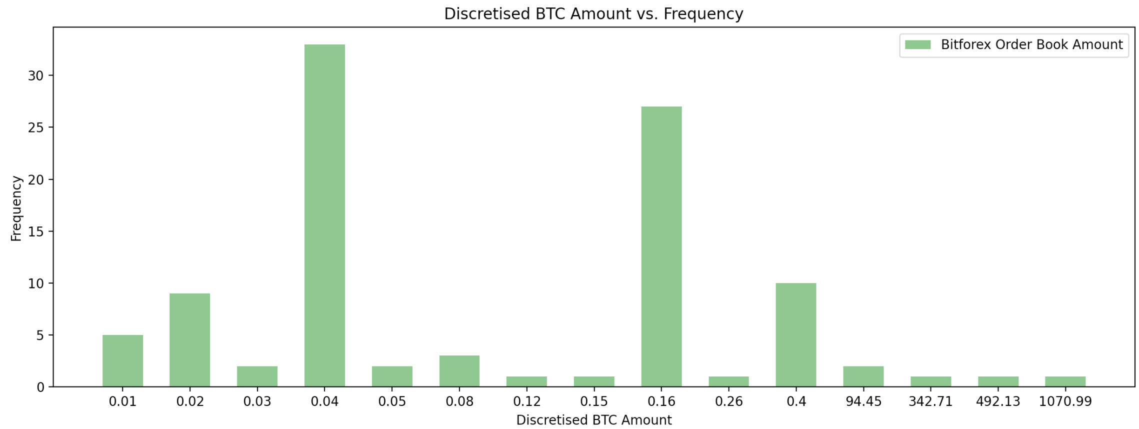
For rounded values, or values that cleanly end in 2dp or less such as 0.55BTC, we can create a bar plot for each exchange. As confirms with our intuition, Coinbase had a large number of 0.1BTC transactions (roughly 1% of their total count). In total, 782 transactions, or 15.6% of all transactions, were in 2dp or fewer.



Coinbase

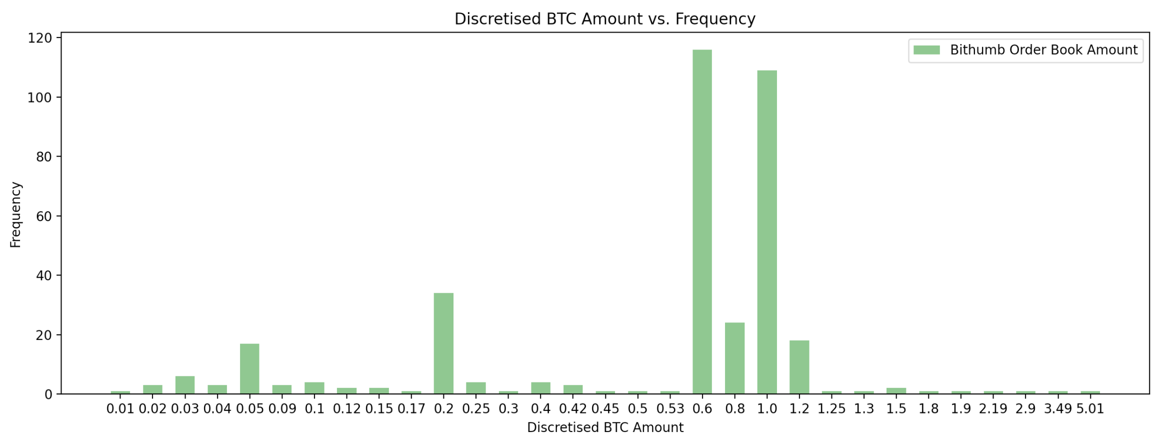
Most other exchanges demonstrate highly suspicious behaviour, with some serving as definitive proof of wash trading.

Bitforex only has 68 transactions out of 5000, or 1.36%, that were 2dp or fewer, or less than a tenth of Coinbase's ratio.



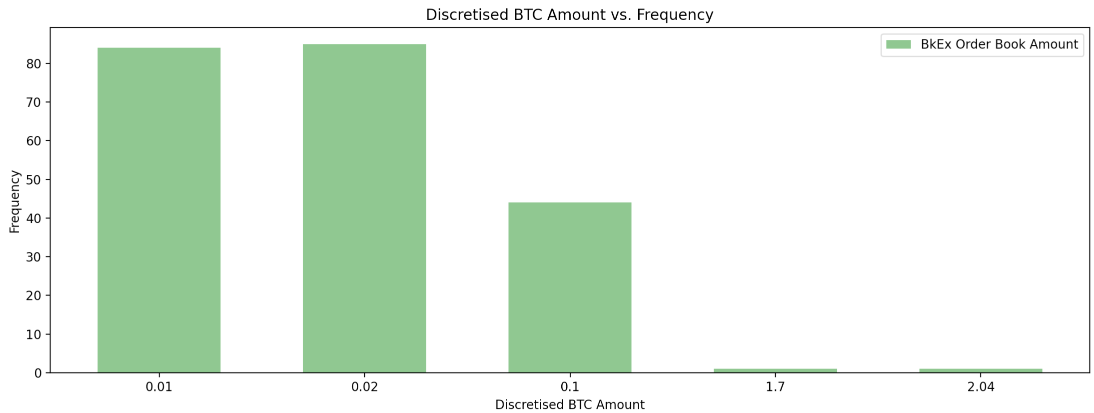
Bitforex

For Bithumb, since there were 340 transactions that were rounded, around half of Coinbase, there is not enough information to warrant clear suspicion.



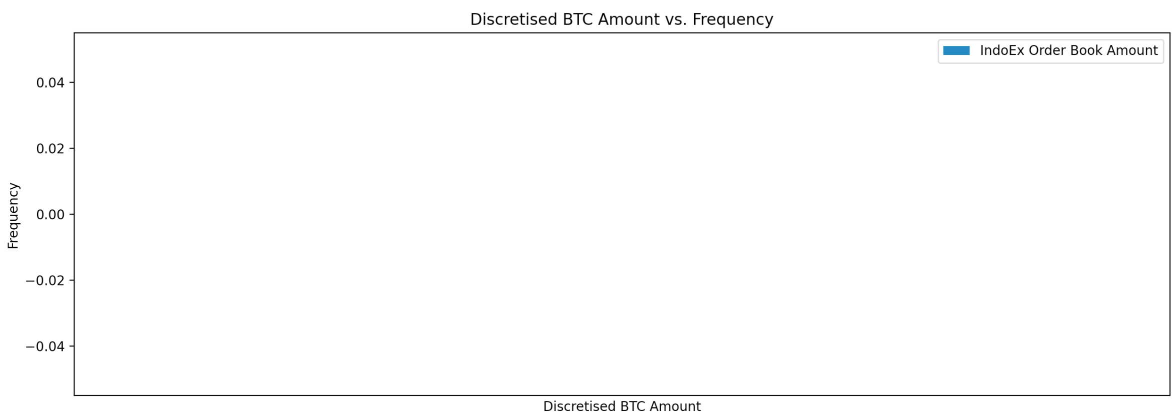
Bithumb

While there were 215 transactions for BkEx, the distribution of them is clearly suspect. They are uniformly concentrated in 0.01 and 0.02, suggesting deliberate manipulation.



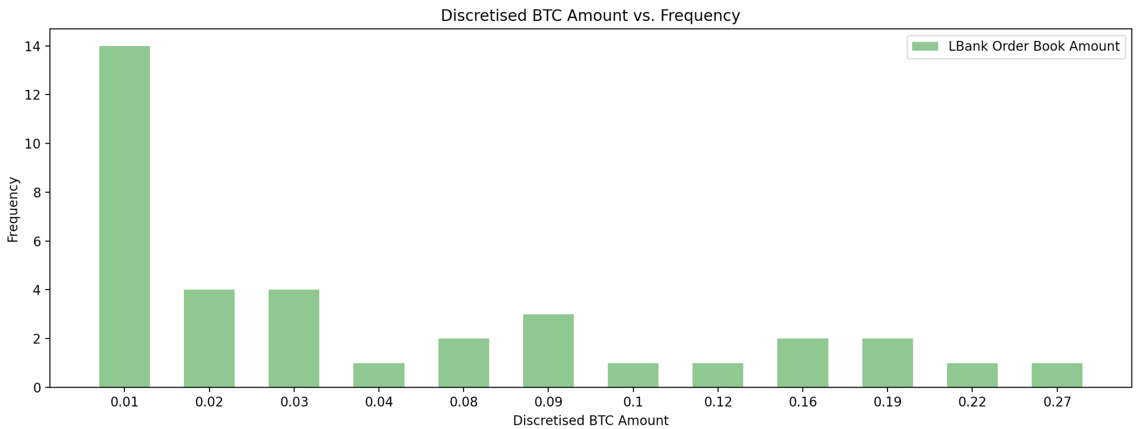
BkEx

There are simply 0 rounded transactions on IndoEx. This is potentially a sign of 0 real customer activity.



IndoEx

LBank only has a tiny 28 transactions that are rounded, or 0.6% of their 5000 transactions collected, be rounded to 2dp which is again, highly suspicious.



LBank

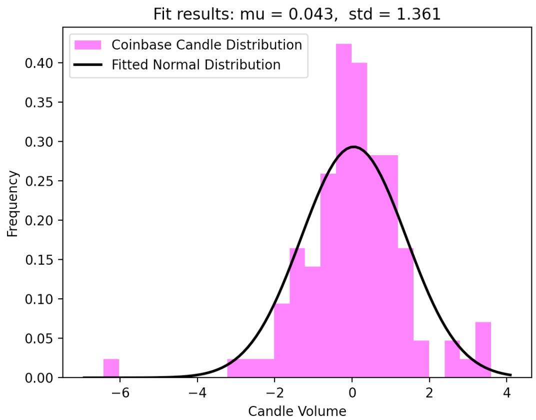
Overall, the rounding raises extremely high suspicion for IndoEx, and high suspicion for LBank, BkEx, and Bitforex.

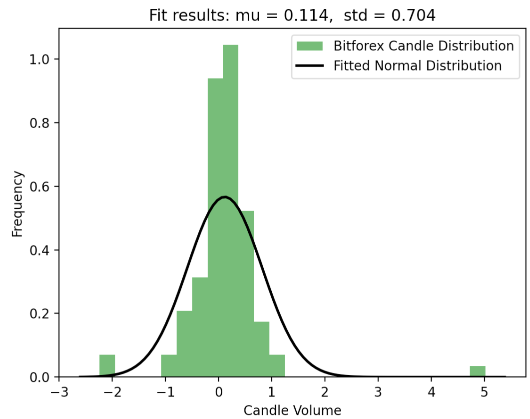
Heuristic 3: Candle Size

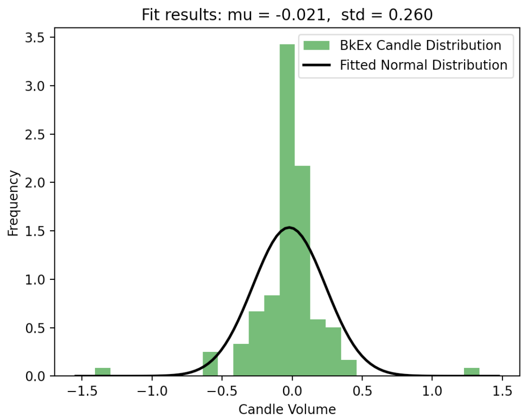
Fitting the normal of Coinbase using MLE gives us the following as the optimal μ and σ^2values, which we can find for every other exchange. Ranking their proximity to the fitted values for coinbase:

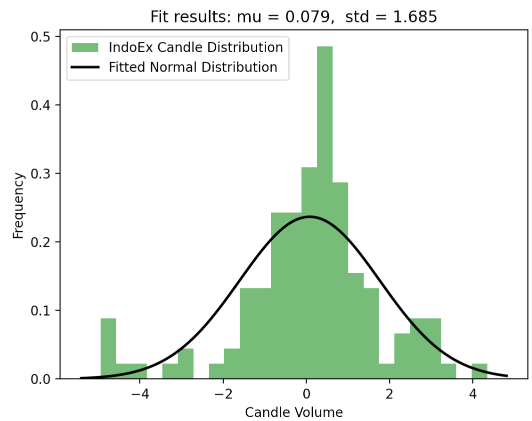
Exchanges	**Fitted Values (optimal μ and σ^2)**	**'Distance' to Coinbase, i.e $\sqrt{(\mu_1 - \mu_2)^2 + (\sigma_1^2 - \sigma_2^2)^2}$**
Coinbase	[0.03998187, 1.36001449]	0
Bitforex	[0.04328145, 1.36098091]	0.0034
IndoEx	[0.07868139, 1.68467228]	0.3270
Bithumb	[0.12040285, 0.558623]	0.8054
BkEx	[-0.02118488, 0.25976902]	1.1019
LBank	[-0.11289942, 13.05990005]	11.7009

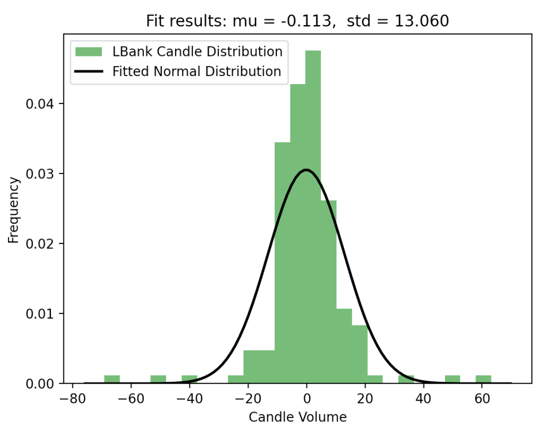
Bithumb, BkEx warrant some suspicion, with LBank having an especially high standard deviation of 13 that is highly suspicious. Both Bithumb and BkEx, on the other hand, have spreads that are unrealistically narrow, lending them to suspicion (between 19% ~ 41% of Coinbase's). This could be a signal that many buys and sells are cancelling out in a single candle; wash trading is happening.

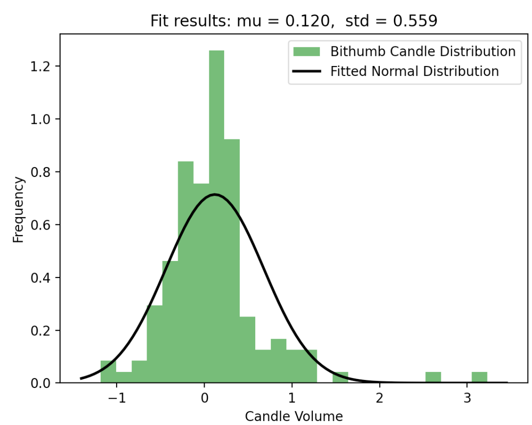












Conclusion:

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

	Intuition 1: Chi-Squared Test Result: x^2 value	Intuition 2.1: Trade Volume Parameter "distance" to Coinbase	Intuition 2.2: Total # of Trade Volume ending in 2dp from 5000	Intuition 3: Candle Volume Parameters "distance" to Coinbase	Cumulatively: Does this exchange wash trade?
Coinbase	🟢 623.098	🟢 0	🟢 782	🟢 0	❌
Bitforex	🟢 747.618	🟡 0.767	🟡 68	🟢 0.003	❌
Bithumb	🟡 2056.713	🟢 0.233	🟢 340	🟢 0.3270	❌
BkEx	🟢 700.629	🛑 2.767	🟡 215	🟡 0.805	✅
IndoEx	🟡 1902.310	🛑 9543.024	🛑 0	🟡 1.102	✅
LBank	🛑 4605.958	🟢 0.233	🛑 28	🛑 11.701	✅

Overall, there was strong evidence of behaviour suggestive of wash trading in these major exchanges. This is disappointing, but also good to know. Given this conclusion, I will finish with improvements that can be made to the model and future areas of inquiry that would be interesting and valuable. There are three main improvements that could be made to the model.

i) Additional Vectors of Analysis

a) 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.

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

ii) 'From scratch' statistical model

It is better to make a more rigorous approach by starting from first principles and build a statistical model, which would inform the kinds of distributions I can expect. However, trade sizing is a tricky model to create from scratch, so I chose to fit a distribution instead. Also, collecting and processing the data was too time consuming to do much else.

iii) *Promising areas of adjacent future research*

a) Inference on specific amount of wash trading

Through this information, we've only arrived at a good idea of what wash trading does, or does 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 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.

Appendix:

Method:

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:

1. 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.*

2. 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.

3. IID assumption in intuition #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, $\frac{p'}{p} \neq 1$, 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 and at higher volume. Constructing a demand & supply model could help with this.

4. 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 into buys, which assumed high conviction in potential buys actually being buys, and likewise for sell.