Accumulation Opportunity

February 2, 2025

1 Introduction

Here we assess the feasibility of accumulating large positions while attempting to maintain low trading costs in an electronic market, in this case a few days per year, including some data from 2024.

2 Data

Load the high-frequency data for all the pairs from the class website. The trade and book data has the following structure

2.1 Format

The data has the following structure[1]

2.1.1 Trades

received_utc_nanoseconds	timestamp_utc_nanoseconds	PriceMilliionths	SizeBilliionths	Side
1674521267814309000	1674521267874527000	22970120000	87069600	-1
1674521267814046000	1674521267874527000	22970150000	25797600	-1
1674611962312088000	1674611962347434000	22499070000	4801640	-1
1674611962339264000	1674611962375191000	22498910000	1120200	-1

The *Side* is actually a sum of trade sides at the same price and time.

2.1.2 Book

Ask1PriceMilliionths	22972550000	22972550000	22502670000	22502670000
Bid1PriceMilliionths	22970150000	22970150000	22498910000	22498910000
Ask1SizeBilliionths	210000000	410000000	101856140	101856140
Bid1SizeBilliionths	25797600	25797600	280050	280050
Ask2PriceMilliionths	22972560000	22972560000	22502680000	22502680000
Bid2PriceMilliionths	22970120000	22970120000	22498690000	22498690000
Ask2SizeBilliionths	210000000	210000000	50000000	50000000
Bid2SizeBilliionths	87069610	87069610	12560150	12560150
received_utc_nanoseconds	1674521267750919800	1674521267751154000	1674611962359972000	1674611962365237000
timestamp_utc_nanoseconds	1674521267806932000	1674521267807073000	1674611962398574000	1674611962400579000
Mid	22971350000	22971350000	22500790000	22500790000

(transposed)

Note: only the sign of the *Side* is important to us. Notice also how in the trade book we have a repeated timestamp, indicating this seller blew through about 9 hundredths at the top level and into close to 3 hundredths at the next level.

[1] Note that inaccuracies in clock settings, i.e. "clock skew", can cause timestamps to appear later than the time at which they are recorded as having been received.

3 Exercise

Write a simple VWAP participation algorithm that takes a target quantity Q, start time τ_s (also known as *arrival time*) and target participation rate p as parameters, and simulates the accumulation possible for a post-only (passive) VWAP algorithm starting at time τ_s, both for buying (positive Q) and for selling (negative Q) in one of the cryptocurrency market data sets provided.

One way to choose a couple values of Q is to compute quantiles of, say, 5 minute volumes. You can then select Q to be a small fraction of, say, the 65th percentile.

Compute or choose a "minimum size threshold" g representing your unlikelihood of being in the primary queue position at any given level. A decent choice for this can come from taking the 5th percentile of trade sizes across the pair's data set.

Your simulation function should use actual signed trading volumes to judge simulated accumulation. It needs to select a quoting participation rate $k_{\text{Pair}}(p)$ that, by necessity, is larger than p due to some trades being unavailable for participation. k should never exceed 5%.Your algorithms should make the conservative assumptions. Namely

- For positive (negative) Q the "available" trades are the ones that were aggressively traded against it, which is to say the ones with negative (positive) *Side*.
- For rach timestamp with trades, sum up all the simultaneous trades' quantities together at each price level to a total q_{level}. The size to assume that you accumulate in any level is $s = \max(0, (q - g)k)$.
- In a flurry of trades trading all the way through one or more available levels[2], you participate in all levels using the formula above, and then no more trades for a pause of P seconds, where P is in the range 0.05 to 5.

Usually we allow k to be a little 'too large" and then allow the algorithm to take pauses when it gets significantly ahead of its target rate p during its runtime.

Assume transaction fees are 50 basis points (0.5% of traded notional) for transactions between crypto-tokens and traditional currencies, and are 10 basis points for trades between cryptotokens.

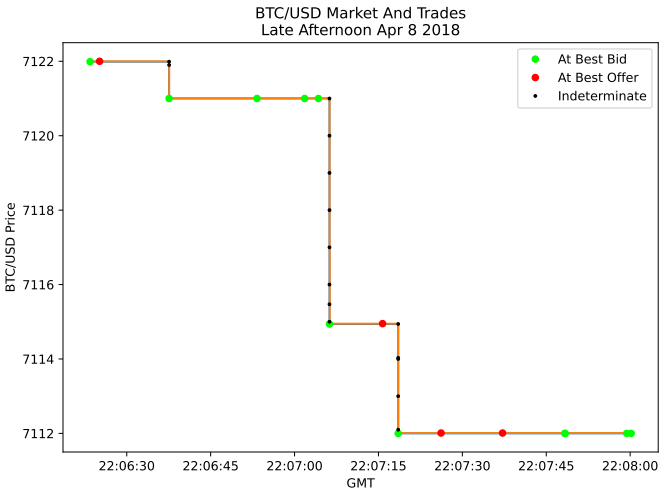

Figure 1: Cases of large sell orders crossing the spread, taking out several levels of resting buy orders.

[2] As, for example, when a large trade "takes out" several levels of the book.

4 Analysis

Assess and contrast the accumulation opportunities available in cryptotoken markets in 2022, 2023 and 2024. You need not check both sides of token-token pairs. Concentrate on transaction sets that finish in 1-15 minutes or so, in order to give yourself a nice set of independent instances for each pair, for which you can get histograms of participation rate accuracy and of average price relative to VWAP and to arrival price. Contrast with the *actual* VWAP in the same periods. Characterize good and bad choices for k.

You can consider metrics such as $\frac{\text{Notional}}{\text{Time}}$, trading costs as a proportion of notional, statistical metrics comparing average price of accumulated positions to arrival prices, relationship between Q and the likelihood of completing the entire quantity in some target wall-clock time horizon, and/or deviation from expected accumulation rate.