

Title: Preliminary Evidence that Retail Trades can be Identified and Counted on the Tape

Author: imthawalrus

Created 2021-10-21 17:40:49 UTC

Permalink: /r/DDintoGME/comments/qcx5dw/preliminary_evidence_that_retail_trades_can_be/

Url: https://www.reddit.com/r/DDintoGME/comments/qcx5dw/preliminary_evidence_that_retail_trades_can_be/

Using the 'Buy' volume shown in half hour intervals in the SEC report just released (Figure 6), I estimated the volume per bar with pixel approximation to graph out total buy volume per 30 minutes between 1/19/21 and 2/5/21. In an attempt to match the volume from the SEC to volume from trades in those intervals, I then downloaded (and cleaned) all trades in the Time and Sales data between those dates. Using a clustering algorithm that adheres to minimum cluster sizes with trade volume as weights, I experimented with the first 30 minutes of trades with the first volume bar from the SEC report as a minimum cluster size to see if we could easily sort out which trades the SEC counts as 'buy' volume (which, since HFT and MM 'buy' volume was excluded, should be all retail 'buy' volume). The results were a bit surprising but very promising because when mapped out by subpenny price, any trades priced over \$XX.XX1000 appear to be retail buys:

[Volume Clustered over Subpenny Prices](<https://imgur.com/a/A3BnozW>)

This shows the volume that the clustering algorithm labeled as retail buys in red vs all other volume in blue and the total volume of red bars equals the volume of the first 'Buy' volume bar in [the SEC report Figure 6](<https://i.redd.it/ykkwpro9w9u71.png>). The numbers across the bottom are the subpenny price ranges that the trades transacted at with midpoint marked as '5000!' (I've expanded on the importance of subpenny priced trades in previous posts).

This is especially interesting according to the research I've done up to this point, internalizers will typically keep retail buys above the subpenny midpoint (\$XX.XX5000) since that allows the internalizer to keep more of the penny fraction but it looks like they were willing to give up that profit to keep retail buys from driving up the NBBO. My next step here is to try to cluster trades for the remaining half hour intervals from the report to come up with a set of training data for a binary classifier to count retail trades outside that 2 week period (but this may take a while since the problem of subset integer addition is NP complete thus the clustering takes quite a while to run).

****TL;DR There's a chance that we can count all retail 'buys' on the tape and come up with a running total to show how much of the float is held by retail traders.****