



# WRDS INTRADAY INDICATOR DATA

---

**Millisecond IID V2.0  
User Manual**

# Contents

|  |        |
|--|--------|
| Executive Summary                        | Page3  |
| Data Content                             | Page4  |
| Data Variable List                       | Page5  |
| Formulas                                 | Page24 |
| FAQ                                      | Page33 |
| Release Notes and In-development Updates | Page40 |

# Executive Summary

The NYSE Trade and Quote (TAQ) database contains intraday transactions data (trades and quotes) for all securities listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), the Nasdaq National Market System (NMS), and all other U.S. equity exchanges.

TAQ is the primary database for intraday market research. With TAQ, a researcher can perform such analyses as daily volatility estimation, probability of informed trading, short-term impact of breaking news, back-testing of intraday trading strategies, and assessment of programmed trading protocols.

TAQ is also the only WRDS database that records market information at the intraday level (to the microsecond), covering roughly 8,000 stock issues listed on major American exchanges. Because it provides a record of every trade and quote on these exchanges (since January 1993), it is by far the largest WRDS database, growing about 0.75 terabytes monthly. No other database has market information at this level of detail.

WRDS acknowledge the computation challenges with TAQ data (the gigantic size of TAQ data - with over 200 TB in total data size; and over 8,000 individual data files). We strive to help TAQ users and market microstructure research. WRDS utilizes its computation power to do the heavy-duty computing work and provides users with a simple and easy-to-use solution. WRDS follow the most recent market microstructure research and construct efficient codes to compute key variables in TAQ research. We provide daily summary with 198 intraday variables covering prices, flows, trade classification, market summaries, liquidity, volatility and informativeness measures covering the entire TAQ sample period. WRDS also includes the interim output such as the complete NBBO and consolidated trades so that users can create their own mythology and variables for research innovations. This suite is named as "WRDS Intraday Indicator Dataset (IID)".

In this user manual, we will provide an overview of WRDS complete NBBO file, WRDS consolidated trades, and the details of the 198 IID daily summary variables. We also list the corresponding formulas in the formula section of this manual.

# Data Content

- I. WRDS Intraday Indicator (Millisecond timestamp, 2003-present)
  - A. WRDS Complete National Best Bid & Offer (NBBO)
 

We add quote observations that have both the national best bid and national best ask to the NBBO files from NYSE.
  - B. WRDS Consolidated Trades (WCT)
 

We match the trades with their prevailing national best bid and offer prices based on the trade and quote timestamp; Various timestamp updates from millisecond to microsecond, from microsecond to nanosecond in both CTA and UTP tapes are carefully handled.
  - C. WRDS Intraday Indicator Dataset (IID) - Daily Summary
    - (i) [Symbol Headers and Time Summaries](#)
    - (ii) [Market Hour \(Open to Close\) Summaries](#)
    - (iii) [After Market Hour \(Before Open/After Close\) Summaries](#)
    - (iv) [Lee and Ready Trade Classification and Liquidity Measures](#)
    - (v) [Tick Test Trade Classification and Summaries](#)
    - (vi) [WRDS Trade Classification and Summaries](#)
    - (vii) [Special Order Summaries](#)
    - (viii) [Volatility and Informativeness Measures](#)
    - (ix) [Retail and Institutional Order Flow](#)
  - D. Update schedule
 

Complete NBBO, WCT, and daily summary of millisecond IID are expected to update daily.
  
- II. WRDS Intraday Indicator (Second timestamp, 1993-2014)
  - A. WRDS NBBO
  - B. WRDS Consolidated Trades (WCT)
  - C. WRDS IID Daily Summary
  - D. Update schedule
 

No update is needed as monthly TAQ is discontinued by NYSE.
  
- III. Millisecond IID Unix location on WRDS server:
 

/wrds/nyse/sasdata/wrds\_taqms\_iid

## Data Variable List (Millisecond IID)

### A. WRDS Complete NBBO

- We fill the official NBBO file from NYSE with the quotes that contains both national best bid and national best offer from the quote file. Quotes with NBB and NBO on both sides are not included in the official NBBO file. We identify these quotes by sorting on quote NBBO conditions:
  - UTP quotes: Qu\_Source = "N" and NatBBO\_Ind='4'
    - 4 = This quote is itself the new NBBO. No appendage is required
  - CTA quotes: Qu\_Source = "C" and NatBBO\_Ind in ('1', 'G')
    - 1= Quote contains Best Bid, Quote contains Best Offer. New quote is itself the Best Bid and Best Offer. No appendage is required. (From 2017 October, NYSE changed this NBBO indicator from numeric value '1' to alphabetic value 'G')
- WRDS Complete NBBO files are indexed by sym\_root and sym\_root-sym\_suffix pairs.
- The complete NBBO file does not eliminate after market-hour quotes. We left them as they were in the NYSE NBBO files. Researchers would expect to see quotes from 4 am to 10 pm.

### B. WRDS Consolidated Trades (WCT)

- We match trades with prevailing quotes according to the timestamp of trades and quotes. We notice the timestamp changed from millisecond to microsecond, and from microsecond to nanosecond from 2003 to 2018.

| DATE OF CHANGE     | UTP CHANGES                         | CTA CHANGES                         |
|--------------------|-------------------------------------|-------------------------------------|
| January 1993       | <b>seconds</b> HHMMSS               | <b>seconds</b> HHMMSS               |
| October 2003       | <b>milliseconds</b> HHMMSSxxx       | <b>milliseconds</b> HHMMSSxxx       |
| July 27, 2015      | <b>microseconds</b> HHMMSSxxxxxx    |                                     |
| August 3, 2015     |                                     | <b>microseconds</b> HHMMSSxxxxxx    |
| October 24, 2016   | <b>nanoseconds</b> HHMMSSxxxxxxxxxx | <b>microseconds</b> HHMMSSxxxxxx000 |
| September 18, 2017 |                                     | <b>nanoseconds</b> HHMMSSxxxxxxxxxx |

We match the trades with previous millisecond/microsecond/nanosecond according to the UTP and CTA timestamp changes listed above.

- WCT files are indexed by sym\_root and sym\_root-sym\_suffix pairs.

## C. WRDS IID Daily Summary

I. Symbol Headers and Time Summaries

| Num | Variable         | Descriptions   | Clarification  | Format        | Example1   | Example 2  |
|-----|------------------|--|--|---------------|------------|------------|
| 1   | DATE             | Date   |  | YYMMDD<br>N8. | 20180925   | 20180925   |
| 2   | SYM_ROOT         | Security symbol root   |  | Char          | AAPL       | IBM        |
| 3   | SYM_SUFFIX       | Security symbol<br>suffix  |  | Char          |            |            |
| 4   | Symbol           | Stock Symbol   | Sym_root+<br>Sym_suffix  | Char          | AAPL       | IBM        |
| 5   | Otime            | Opening Time<br>(First Market Center<br>Official Opening<br>Trade)                           | Trade<br>Condition<br>= 'Q' or 'O';<br>When<br>multiple<br>market<br>centers<br>report<br>opening<br>trades, the<br>first market<br>center is<br>recorded;<br><br>When both<br>'Q' and 'O'<br>are<br>reported,<br>we use 'O' | TIME18.9      | 9:30:00 AM | 9:30:01 AM |
| 6   | OEX              | The First Opening<br>Exchange Market<br>Center (with the First<br>Official Opening<br>Trade) |  | Char          | P          | P          |
| 7   | Oprc             | Opening Price<br>(First Market Center<br>Official Opening<br>Trade)                          |  | BEST32.       | 219.76     | 150.72     |
| 8   | Osize            | Opening Trade<br>Volume<br>(First Market Center<br>Official Opening<br>Trade)                |  | BEST32.       | 78         | 30         |
| 9   | CTime            | Closing Time<br>(Last Market Center<br>Official Close Trade)                                 | Trade<br>Condition<br>= 'M' or '6';<br>When<br>multiple<br>market<br>centers<br>report<br>closing<br>trades, the<br>last market<br>center is<br>recorded;<br><br>When both<br>'M' and '6'<br>are<br>reported,<br>we use '6'  | TIME18.9      | 4:00:00 PM | 4:00:00 PM |
| 10  | CEX              | The Last Closing<br>Exchange Market<br>Center<br>(with the Last Official<br>Close Trade)     |  | Char          | P          | P          |
| 11  | CPrc             | Closing Price<br>(Last Market Center<br>Official Close Trade)                                |  | BEST32.       | 222.19     | 148.91     |
| 12  | Csize            | Closing Trade<br>Volume<br>(Last Market Center<br>Official Close Trade)                      |  | BEST32.       | 500        | 2400       |
| 13  | nbbot_after_open | Best Bid/Offer Quote<br>Time of the Most   |  | TIME18.9      | 9:30:00 AM | 9:30:01 AM |

|    |                   | Recent Quote since Market Open   |                                      |          |            |             |
|----|-------------------|--|--------------------------------------|----------|------------|-------------|
| 14 | mid_after_open    | Best Bid/Offer Middle Price of the Most Recent Quote since Market Open | mid price<br>=(best bid+ best ask)/2 | BEST32.  | 219.78     | 150.09      |
| 15 | ttime_open        | Trade Time of the Most Recent Trade since Market Open                  |                                      | TIME18.9 | 9:30:00 AM | 9:30:02 AM  |
| 16 | stime_open        | Trade Volume of the Most Recent Trade since Market Open                |                                      | BEST32.  | 31         | 100         |
| 17 | ptime_open        | Trade Price of the Most Recent Trade since Market Open                 |                                      | BEST32.  | 219.799    | 149.91      |
| 18 | nbb_after_open    | Best Bid Price of the Most Recent Quote since Market Open              |                                      | BEST32.  | 219.76     | 149.54      |
| 19 | nbbqty_after_open | Best Bid Size of the Most Recent Quote since Market Open               |                                      | BEST32.  | 200        | 300         |
| 20 | nbo_after_open    | Best Offer Price of the Most Recent Quote since Market Open            |                                      | BEST32.  | 219.8      | 150.64      |
| 21 | nboqty_after_open | Best Offer Size of the Most Recent Quote since Market Open             |                                      | BEST32.  | 700        | 100         |
| 22 | nbbot_1pm         | Most Recent Best Bid/Offer Quote Time as of 1PM                        |                                      | TIME18.9 | 1:00:00 PM | 12:59:58 PM |
| 23 | mid_1pm           | Most Recent Best Bid/Offer Middle Price as of 1PM                      |                                      | BEST32.  | 222.09     | 149.035     |
| 24 | ttime_1pm         | Trade Time of the Most Recent Trade after 1 pm                         |                                      | TIME18.9 | 1:00:00 PM | 1:00:02 PM  |
| 25 | ptime_1pm         | Trade Price of the Most Recent Trade after 1 pm                        |                                      | BEST32.  | 222.1      | 149.03      |
| 26 | size_1pm          | Trade Volume of the Most Recent Trade after 1 pm                       |                                      | BEST32.  | 20         | 100         |
| 27 | nbb_1pm           | Most Recent Best Bid Price as of 1PM                                   |                                      | BEST32.  | 222.08     | 149.03      |
| 28 | nbbqty_1pm        | Most Recent Best Bid Size as of 1PM                                    |                                      | BEST32.  | 200        | 300         |
| 29 | nbo_1pm           | Most Recent Best Offer Price as of 1PM                                 |                                      | BEST32.  | 222.1      | 149.04      |

|    |                     |   |   |          |            |            |
|----|---------------------|---|---|----------|------------|------------|
| 30 | nboqty_1pm          | Most Recent Best Offer Size as of 1PM                             |   | BEST32.  | 100        | 300        |
| 31 | nbbot_4pm           | Last Best Bid/Offer Quote Time before 4PM                         |   | TIME18.9 | 4:00:00 PM | 4:00:00 PM |
| 32 | mid_4pm             | Last Best Bid/Offer Middle Price before 4PM                       |   | BEST32.  | 222.195    | 148.915    |
| 33 | ttime_4pm           | Trade Time of the Most Recent Trade before 4 pm                   |   | TIME18.9 | 4:00:00 PM | 4:00:00 PM |
| 34 | ptime_4pm           | Trade Price of the Most Recent Trade before 4 pm                  |   | BEST32.  | 222.19     | 148.91     |
| 35 | size_4pm            | Trade Volume of the Most Recent Trade before 4 pm                 |   | BEST32.  | 500        | 95         |
| 36 | nbb_4pm             | Last Best Bid Price before 4PM                                    |   | BEST32.  | 222.19     | 148.91     |
| 37 | nbbqty_4pm          | Last Best Bid Size before 4PM                                     |   | BEST32.  | 100        | 1700       |
| 38 | nbo_4pm             | Last Best Offer Price before 4PM                                  |   | BEST32.  | 222.2      | 148.92     |
| 39 | nboqty_4pm          | Last Best Offer Size before 4PM                                   |   | BEST32.  | 7700       | 2300       |
| 40 | nbbot_before_close  | Best Bid/Offer Quote Time of the Last Quote before Market Close   |   | TIME18.9 | 4:00:00 PM | 4:00:00 PM |
| 41 | mid_before_close    | Best Bid/Offer Middle Price of the Last Quote before Market Close | mid price<br>=(best bid+<br>best ask)/2 | BEST32.  | 222.195    | 148.915    |
| 42 | ttime_close         | Trade Time of the Last Trade before Market Close                  |   | TIME18.9 | 4:00:00 PM | 4:00:00 PM |
| 43 | ptime_close         | Trade Price of the Last Trade before Market Close                 |   | BEST32.  | 222.19     | 148.92     |
| 44 | stime_close         | Trade Volume of the Last Trade before Market Close                |   | BEST32.  | 500        | 100        |
| 45 | nbb_before_close    | Best Bid Price of the Last Quote before Market Close              |   | BEST32.  | 222.19     | 148.91     |
| 46 | nbbqty_before_close | Best Bid Size of the Last Quote before Market Close               |   | BEST32.  | 100        | 1700       |



|     |                     |   |  |         |            |           |
|-----|---------------------|---|--|---------|------------|-----------|
| 47  | nbo_before_close    | Best Offer Price of the Last Quote before Market Close  |  | BEST32. | 222.2      | 148.92    |
| 48  | nboqty_before_close | Best Offer Size of the Last Quote before Market Close   |  | BEST32. | 7700       | 2300      |
| 49  | avg_price_m         | Average Trade Price during market hours (Open to Close)   | Requires the stock to have at least one official opening trade and one official closing trade; otherwise use 9:30 am to 4 pm | BEST32. | 221.562    | 149.251   |
| 50  | price_high_m        | Highest Trade Price during market hours (Open to Close)   |  | BEST32. | 222.82     | 150.72    |
| 51  | price_low_m         | Lowest Trade Price during market hours (Open to Close)  |  | BEST32. | 201.5      | 146.028   |
| 52  | total_dollar_m      | Total Trade Value in Dollars during market hours (Open to Close)                                    |  | BEST32. | 4207411930 | 474040084 |
| 53  | total_n_trades_m    | Total Number of Trades during market hours (Open to Close)  |  | BEST32. | 169441     | 32357     |
| 54  | total_vol_m         | Total Trade Volume during market hours (Open to Close)  |  | BEST32. | 18993554   | 3175237   |
| 55  | vw_price_m          | Volume-weighted Average Trade Price during market hours (Open to Close)                             |  | BEST32. | 221.518    | 149.293   |
| 156 | TR_SEQNUM_open      | Trade Sequence Number of First Official Open Trade  |  | Num     | 2816       | 4788      |
| 157 | TR_SEQNUM_close     | Trade Sequence Number of the Last Official Close Trade  |  | Num     | 1288422    | 1784642   |
| 158 | O_official          | Indicator for Official Opening Trade Flag (Use First Trade after 9 pm if No Official Opening Trade) | ('1' if official opening trade exists, '0' otherwise)  | Num     | 1          | 1         |
| 159 | C_official          | Indicator for Official Closing Trade Flag (Use Last Trade before 4 pm if No Official Closing Trade) | ('1' if official closing trade exists, '0' otherwise)  | Num     | 1          | 1         |

## II. Market Hour (Open to Close) Summaries

| Num | Variable                | Descriptions  | Clarification | Format  | Example     | Example 2  |
|-----|-------------------------|---|---------------|---------|-------------|------------|
| 49  | avg_price_m             | Average Trade Price during market hours (Open to Close)                 |               | BEST32. | 221.562     | 149.251    |
| 50  | price_high_m            | Highest Trade Price during market hours (Open to Close)                 |               | BEST32. | 222.82      | 150.72     |
| 51  | price_low_m             | Lowest Trade Price during market hours (Open to Close)                  |               | BEST32. | 201.5       | 146.028    |
| 52  | total_dollar_m          | Total Trade Value in Dollars during market hours (Open to Close)        |               | BEST32. | 4207411930  | 474040084  |
| 53  | total_n_trades_m        | Total Number of Trades during market hours (Open to Close)              |               | BEST32. | 169441      | 32357      |
| 54  | total_vol_m             | Total Trade Volume during market hours (Open to Close)                  |               | BEST32. | 18993554    | 3175237    |
| 55  | vw_price_m              | Volume-weighted Average Trade Price during market hours (Open to Close) |               | BEST32. | 221.518     | 149.293    |
| 56  | BestBidDepth_Dollar_tw  | Time-Weighted Best Bid Dollar Depth (During Market Hours)               |               | BEST32. | 65909.92    | 42701.33   |
| 57  | BestBidDepth_Share_tw   | Time-Weighted Best Bid Share Depth (During Market Hours)                |               | BEST32. | 297.15      | 286.213    |
| 58  | BestOfrDepth_Dollar_tw  | Time-Weighted Best Offer Dollar Depth (During Market Hours)             |               | BEST32. | 65278.84    | 37673.36   |
| 59  | BestOfrDepth_Share_tw   | Time-Weighted Best Offer Share Depth (During Market Hours)              |               | BEST32. | 294.25      | 252.463    |
| 60  | QuotedSpread_Dollar_tw  | Time-Weighted Dollar Quoted Spread (During Market Hours)                |               | BEST32. | 0.018018    | 0.021347   |
| 61  | QuotedSpread_Percent_tw | Time-Weighted Percent Quoted Spread (During Market Hours)               |               | BEST32. | 0.000081311 | 0.00014277 |

### III. After Market Hour (Before Open/After Close) Summaries

| Num | Variable         | Descriptions  | Clarification | Format  | Example    | Example 2 |
|-----|------------------|---|---------------|---------|------------|-----------|
| 129 | avg_price_b      | Average Trade Price outside market hours (Before Market Open)                 |               | BEST32. | 221.562    | 149.251   |
| 130 | price_high_b     | Highest Trade Price outside market hours (Before Market Open)                 |               | BEST32. | 222.82     | 150.72    |
| 131 | price_low_b      | Lowest Trade Price outside market hours (Before Market Open)                  |               | BEST32. | 201.5      | 146.028   |
| 132 | total_dollar_b   | Total Trade Value in Dollars outside market hours (Before Market Open)        |               | BEST32. | 4207411930 | 474040084 |
| 133 | total_n_trades_b | Total Number of Trades outside market hours (Before Market Open)              |               | BEST32. | 169441     | 32357     |
| 134 | total_vol_b      | Total Trade Volume outside market hours (Before Market Open)                  |               | BEST32. | 18993554   | 3175237   |
| 135 | vw_price_b       | Volume-weighted Average Trade Price outside market hours (Before Market Open) |               | BEST32. | 221.518    | 149.293   |
| 136 | avg_price_a      | Average Trade Price outside market hours (After Market Close)                 |               | BEST32. | 221.562    | 149.251   |
| 137 | price_high_a     | Highest Trade Price outside market hours (After Market Close)                 |               | BEST32. | 222.82     | 150.72    |
| 138 | price_low_a      | Lowest Trade Price outside market hours (After Market Close)                  |               | BEST32. | 201.5      | 146.028   |
| 139 | total_dollar_a   | Total Trade Value in Dollars outside market hours (After Market Close)        |               | BEST32. | 4207411930 | 474040084 |
| 140 | total_n_trades_a | Total Number of Trades outside market hours (After Market Close)              |               | BEST32. | 169441     | 32357     |
| 141 | total_vol_a      | Total Trade Volume outside market hours (After Market Close)                  |               | BEST32. | 18993554   | 3175237   |
| 142 | vw_price_a       | Volume-weighted Average Trade Price outside market hours (After Market Close) |               | BEST32. | 221.518    | 149.293   |

#### IV. Lee and Ready Trade Classification and Liquidity Measures

| Num | Variable                    | Descriptions   | Clarification   | Format  | Example     | Example 2  |
|-----|-----------------------------|--|---|---------|-------------|------------|
| 62  | BuyNumTrades_LR             | Total Number of Buys (Lee Ready)                           | See formula note: Formula 1, Lee-Ready previous nanosecond algorithm      | BEST32. | 81593       | 15637      |
| 63  | BuyVol_LR                   | Sum of Buy Trade Volume(Lee Ready)                         |   | BEST32. | 8980934     | 1585367    |
| 64  | buy_dv_LR                   | Sum of Buy Trade Value in Dollars (Lee Ready)              |   | BEST32. | 1989923284  | 236670973  |
| 65  | SellNumTrades_LR            | Total Number of Sells (Lee Ready)                          |   | BEST32. | 87848       | 16719      |
| 66  | SellVol_LR                  | Sum of Sell Trade Volume(Lee Ready)                        |   | BEST32. | 10012198    | 1587400    |
| 67  | sell_dv_LR                  | Sum of Sell Trade Value in Dollars (Lee Ready)             |   | BEST32. | 2217394692  | 237001357  |
| 68  | total_trade                 | Total Number of Trades (Lee Ready)                         |   | BEST32. | 169441      | 32356      |
| 69  | total_vol                   | Sum of all trade volume (Lee Ready)                        |   | BEST32. | 18993132    | 3172767    |
| 70  | total_dv_LR                 | Total Trade Value in Dollars(Lee Ready)                    |   | BEST32. | 4207317976  | 473672330  |
| 71  | avg_buy_price_LR            | Average Trade Price of Buys (Lee Ready)                    |   | BEST32. | 221.56      | 149.251    |
| 72  | vwavg_buy_price_LR          | Value-weighted Average Trade Price of Buys (Lee Ready)     | See formula note: Formula 9, Value Weighted Average Price of Buy Trades   | BEST32. | 221.572     | 149.285    |
| 73  | avg_sell_price_LR           | Average Trade Price of Sells (Lee Ready)                   |   | BEST32. | 221.564     | 149.252    |
| 74  | vwavg_sell_price_LR         | Value-weighted Average Trade Price of Sells (Lee Ready)    | See formula note: Formula 10, Value Weighted Average Price of Sell Trades | BEST32. | 221.469     | 149.302    |
| 75  | EffectiveSpread_Percent_Ave | Simple Averaged Percent Effective Spread (Lee Ready)       | See formula note: Formula 4, Simple Averaged Percentage Effective Spread  | BEST32. | 0.000068278 | 9.9602E-05 |
| 76  | EffectiveSpread_Percent_DW  | Dollar Value-weighted Percent Effective Spread (Lee Ready) | See formula note: Formula 13, Value Weighted Percentage Effective Spread  | BEST32. | 0.00007327  | 0.00028208 |
| 77  | EffectiveSpread_Percent_SW  | Share Volume-weighted Percent Effective Spread (Lee Ready) | See formula note: Formula 14, Share Weighted Percentage Effective Spread  | BEST32. | 0.000073357 | 0.0002813  |

|    |                                      |   |  |         |             |            |
|----|--------------------------------------|---|--|---------|-------------|------------|
| 78 | PercentPrice<br>Impact_LR_A<br>ve    | Simple Averaged<br>Percent Price Impact<br>(Lee Ready)              | See formula note:<br>Formula 8, Simple<br>Averaged Percentage<br>Price Impact    | BEST32. | 0.000022604 | 5.9603E-05 |
| 79 | PercentPrice<br>Impact_LR_D<br>W     | Dollar Value-<br>weighted Percent<br>Price Impact<br>(Lee Ready)    | See formula note:<br>Formula 21, Value<br>Weighted Percentage<br>Price Impact    | BEST32. | 0.000156963 | 0.00018574 |
| 80 | PercentPrice<br>Impact_LR_S<br>W     | Share Volume-<br>weighted Percent<br>Price Impact<br>(Lee Ready)    | See formula note:<br>Formula 22, Share<br>Weighted Percentage<br>Price Impact    | BEST32. | 0.000157917 | 0.00018521 |
| 81 | PercentReal<br>izedSpread_L<br>R_Ave | Simple Averaged<br>Percent Realized<br>Spread (Lee Ready)           | See formula note:<br>Formula 6, Simple<br>Averaged Percentage<br>Realized Spread | BEST32. | 0.000045674 | 4.0003E-05 |
| 82 | PercentReal<br>izedSpread_L<br>R_DW  | Dollar Value-<br>weighted Percent<br>Realized Spread<br>(Lee Ready) | See formula note:<br>Formula 18, Value<br>Weighted Percentage<br>Realized Spread | BEST32. | -8.3694E-05 | 9.6353E-05 |
| 83 | PercentReal<br>izedSpread_L<br>R_SW  | Share Volume-<br>weighted Percent<br>Realized Spread<br>(Lee Ready) | See formula note:<br>Formula 17, Share<br>Weighted Percentage<br>Realized Spread | BEST32. | -8.4561E-05 | 9.6109E-05 |
| 84 | EffectiveSpre<br>ad_Dollar_A<br>ve   | Simple Averaged<br>Dollar Effective<br>Spread (Lee Ready)           | See formula note:<br>Formula 3, Simple<br>Averaged Dollar<br>Effective Spread    | BEST32. | 0.015102    | 0.014889   |
| 85 | EffectiveSpre<br>ad_Dollar_D<br>W    | Dollar Value-<br>weighted Dollar<br>Effective Spread<br>(Lee Ready) | See formula note:<br>Formula 11, Value<br>Weighted Dollar<br>Effective Spread    | BEST32. | 0.016213    | 0.042239   |
| 86 | EffectiveSpre<br>ad_Dollar_S<br>W    | Share Volume-<br>weighted Dollar<br>Effective Spread<br>(Lee Ready) | See formula note:<br>Formula 12, Share<br>Weighted Dollar<br>Effective Spread    | BEST32. | 0.016232    | 0.042122   |
| 87 | DollarPricel<br>mpact_LR_A<br>ve     | Simple Averaged<br>Dollar Price Impact<br>(Lee Ready)               | See formula note:<br>Formula 7, Simple<br>Averaged Dollar Price<br>Impact        | BEST32. | 0.004994598 | 0.00890121 |
| 88 | DollarPricel<br>mpact_LR_D<br>W      | Dollar Value-<br>weighted Dollar<br>Price Impact<br>(Lee Ready)     | See formula note:<br>Formula 19, Value<br>Weighted Dollar Price<br>Impact        | BEST32. | -0.018382   | 0.014432   |
| 89 | DollarPricel<br>mpact_LR_S<br>W      | Share Volume-<br>weighted Dollar<br>Price Impact<br>(Lee Ready)     | See formula note:<br>Formula 20, Share<br>Weighted Dollar Price<br>Impact        | BEST32. | -0.018572   | 0.014395   |

|           |                                     |  |  |         |          |          |
|-----------|-------------------------------------|--|--|---------|----------|----------|
| <b>90</b> | DollarRealize<br>dSpread_LR_<br>Ave | Simple Averaged<br>Dollar Realized<br>Spread (Lee Ready)           | See formula note:<br>Formula 5, Simple<br>Averaged Dollar<br>Realized Spread | BEST32. | 0.010107 | 0.005988 |
| <b>91</b> | DollarRealize<br>dSpread_LR_<br>DW  | Dollar Value-<br>weighted Dollar<br>Realized Spread<br>(Lee Ready) | See formula note:<br>Formula 16, Value<br>Weighted Dollar<br>Realized Spread | BEST32. | 0.034595 | 0.02781  |
| <b>92</b> | DollarRealize<br>dSpread_LR_<br>SW  | Share Volume-<br>weighted Dollar<br>Realized Spread<br>(Lee Ready) | See formula note:<br>Formula 15, Share<br>Weighted Dollar<br>Realized Spread | BEST32. | 0.034804 | 0.027731 |

## V. Tick Test Trade Classification and Summaries

| Num | Variable              | Descriptions  | Clarification                          | Format  | Example    | Example 2 |
|-----|-----------------------|---|--|---------|------------|-----------|
| 93  | BuyNumTrades_tick     | Total Number of Buys (Tick Test)                        | See formula note: Formula 2, Tick Test | BEST32. | 82176      | 15428     |
| 94  | BuyVol_tick           | Sum of Buy Trade Volume (Tick Test)                     |  | BEST32. | 9540882    | 1566648   |
| 95  | buy_dv_tick           | Sum of Buy Trade Value in Dollars (Tick Test)           |  | BEST32. | 2112992063 | 233886536 |
| 96  | avg_buy_price_tick    | Average Trade Price of Buys (Tick Test)                 |  | BEST32. | 221.561    | 149.26    |
| 97  | vwavg_buy_price_tick  | Value-weighted Average Trade Price of Buys (Tick Test)  |  | BEST32. | 221.467    | 149.291   |
| 98  | SellNumTrades_tick    | Total Number of Sells (Tick Test)                       |  | BEST32. | 87265      | 16928     |
| 99  | SellVol_tick          | Sum of Sell Trade Volume (Tick Test)                    |  | BEST32. | 9452250    | 1606119   |
| 100 | sell_dv_tick          | Sum of Sell Trade Value in Dollars (Tick Test)          |  | BEST32. | 2094325914 | 239785794 |
| 101 | avg_sell_price_tick   | Average Trade Price of Sells (Tick Test)                |  | BEST32. | 221.564    | 149.244   |
| 102 | vwavg_sell_price_tick | Value-weighted Average Trade Price of Sells (Tick Test) |  | BEST32. | 221.569    | 149.295   |
| 103 | total_trade_tick      | Total Number of Trades (Tick Test)                      |  | BEST32. | 169441     | 32356     |
| 104 | total_vol_tick        | Sum of all trade volume (Tick Test)                     |  | BEST32. | 18993132   | 3172767   |
| 105 | total_dv_tick         | Total Trade Value in Dollars (Tick Test)                |  | BEST32. | 4207317976 | 473672330 |

## VI. WRDS Trade Classification and Summaries

| Num | Variable              | Descriptions  | Clarification                                    | Format  | Example    | Example 2 |
|-----|-----------------------|---|--|---------|------------|-----------|
| 143 | BuyNumTrades_wrds     | Total Number of Buys (WRDS Classification)                        | See formula note: Formula 2, WRDS Classification | BEST32. | 82176      | 15428     |
| 144 | BuyVol_wrds           | Sum of Buy Trade Volume (WRDS Classification)                     |  | BEST32. | 9540882    | 1566648   |
| 145 | buy_dv_wrds           | Sum of Buy Trade Value in Dollars (WRDS Classification)           |  | BEST32. | 2112992063 | 233886536 |
| 146 | avg_buy_price_wrds    | Average Trade Price of Buys (WRDS Classification)                 |  | BEST32. | 221.561    | 149.26    |
| 147 | vwavg_buy_price_wrds  | Value-weighted Average Trade Price of Buys (WRDS Classification)  |  | BEST32. | 221.467    | 149.291   |
| 148 | SellNumTrades_wrds    | Total Number of Sells (WRDS Classification)                       |  | BEST32. | 87265      | 16928     |
| 149 | SellVol_wrds          | Sum of Sell Trade Volume (WRDS Classification)                    |  | BEST32. | 9452250    | 1606119   |
| 150 | sell_dv_wrds          | Sum of Sell Trade Value in Dollars (WRDS Classification)          |  | BEST32. | 2094325914 | 239785794 |
| 151 | avg_sell_price_wrds   | Average Trade Price of Sells (WRDS Classification)                |  | BEST32. | 221.564    | 149.244   |
| 152 | vwavg_sell_price_wrds | Value-weighted Average Trade Price of Sells (WRDS Classification) |  | BEST32. | 221.569    | 149.295   |
| 153 | total_trade_wrds      | Total Number of Trades (WRDS Classification)                      |  | BEST32. | 169441     | 32356     |
| 154 | total_vol_wrds        | Sum of all trade volume (WRDS Classification)                     |  | BEST32. | 18993132   | 3172767   |
| 155 | total_dv_wrds         | Total Trade Value in Dollars (WRDS Classification)                |  | BEST32. | 4207317976 | 473672330 |



## VII. Special Order Summaries

| Num | Variable             | Descriptions   | Clarification  | Format  | Example     | Example 2  |
|-----|----------------------|--|--|---------|-------------|------------|
| 110 | n_iso_trade          | Number of Intermarket Sweep Order Trades (During Market Hours)   |  | BEST32. | 65315       | 12737      |
| 111 | iso_dollar           | Sum of Intermarket Sweep Order Trade Dollar Value (During Market Hours)  |  | BEST32. | 1158594631  | 163206291  |
| 112 | iso_vol              | Sum of Intermarket Sweep Order Trade Volume (During Market Hours)  |  | BEST32. | 5228578     | 1093508    |
| 117 | n_oddlot_trade       | Number of Odd Lot Trades (During Market Hours)   |  | BEST32. | 84234       | 13440      |
| 118 | oddlot_dollar        | Sum of Odd Lot Trade Value (Dollar Volume during Market Hours)   |  | BEST32. | 561898333.6 | 63759435.9 |
| 119 | oddlot_vol           | Sum of Odd Lot Trade Volume (During Market Hours)  |  | BEST32. | 2536195     | 427286     |
| 120 | n_outside_nbbo_trade | Number of trades out of the previous nanosecond National Best Bid-Ask range  |  | BEST32. | 9594        | 919        |
| 121 | ret_mkt_m            | Open to Close Return (During Market Hours) (Note: When official market opening trade or official market closing trade are missing, set to missing) | Log return of official opening price over official closing price | BEST32. | -0.010997   | 0.012082   |

## VIII. Volatility and Informativeness Measures

| Num | Variable      | Descriptions  | Clarification  | Format  | Example     | Example 2 |
|-----|---------------|---|--|---------|-------------|-----------|
| 106 | Hindex        | Herfindahl Index calculated across 30-minute Time Units (During Market Hours)   | See formula note: Formula 32, Herfindahl Index   | BEST32. | 0.10505     | 0.09708   |
| 107 | TSigSqrtDVol1 | Lambda (Price Impact Coefficient) with intercept (During Market Hours)          | See formula note: Formula 30, Lambda (Price Impact Coefficient) 1  | BEST32. | 0.000000201 | 5.2E-08   |
| 108 | TSigSqrtDVol2 | Lambda (Price Impact Coefficient) without intercept (During Market Hours)       | See formula note: Formula 31, Lambda (Price Impact Coefficient) 2  | BEST32. | 0.000000172 | 5.7E-08   |
| 109 | n_obs         | Number of Observations used in the Lambda (Price Impact Coefficient) Regression | Number of 5-minute intervals used in the Lambda Regression; 78 5-min intervals on normal days, less freedom of 1 | BEST32. | 87          | 77        |
| 113 | ivol_q        | Quote-based Intraday Volatility (During Market Hours)                           | See formula note: Formula 24, Second-by-second Intraday Volatility (quote based)                                 | BEST32. | 2.07E-10    | 7.89E-10  |
| 114 | ivol_t        | Trade-based Intraday Volatility (During Market Hours)                           | See formula note: Formula 23, Second-by-second Intraday Volatility (trade based)                                 | BEST32. | 0.000000107 | 3.7E-08   |
| 115 | n30_pos       | Number of 30-minute Time Intervals with Trade (During Market Hours)             | Value='13' on regular days, less than '13' for early closure days  | BEST32. | 23          | 13        |
| 116 | n5_pos        | Number of 5-minute Time Intervals with Trade (During Market Hours)              | Value='78' on regular days, less than '78' for early closure days  | BEST32. | 88          | 78        |

|            |              |  |   |         |          |          |
|------------|--------------|--|---|---------|----------|----------|
| <b>122</b> | var_ratio1   | Variance Ratio 1<br>(15-second/3*5-second) (During Market Hours) | See formula note:<br>Formula 25, Variance Ratio 1 (15-second/3*5-second)  | BEST32. | 0.029621 | 0.092919 |
| <b>123</b> | var_ratio2   | Variance Ratio 2<br>(1-min/4*15-second) (During Market Hours)    | See formula note:<br>Formula 26, Variance Ratio 2 (1-min/4*15-second)   | BEST32. | 0.095376 | 0.011701 |
| <b>124</b> | var_ratio3   | Variance Ratio 3<br>(5-min/5*1-min) (During Market Hours)        | See formula note:<br>Formula 27, Variance Ratio 3 (5-min/5*1-min)   | BEST32. | 0.20443  | 0.12154  |
| <b>125</b> | var_ratio4   | Variance Ratio 4<br>(15-min/3*5-min) (During Market Hours)       | See formula note:<br>Formula 28, Variance Ratio 4 (15-min/3*5-min)  | BEST32. | 0.054    | 0.43729  |
| <b>126</b> | var_ratio5   | Variance Ratio 5<br>(30-min/2*15-min) (During Market Hours)      | See formula note:<br>Formula 29, Variance Ratio 5 (30-min/2*15-min)   | BEST32. | 0.30944  | 0.40678  |
| <b>127</b> | bs_ratio_num | Absolute Percent Order Imbalance<br>- Num of Trades              | See formula note:<br>Formula 33; absolute value of [(num. of buys minus num. of sells) over total number of trades] | BEST32. | 0.036916 | 0.03344  |
| <b>128</b> | bs_ratio_vol | Absolute Percent Order Imbalance<br>- Trade Volume               | See formula note:<br>Formula 34; absolute value of [(buy volume minus sell volume) over total trade volume]         | BEST32. | 0.054297 | 0.000641 |

## IX. Retail and Institutional Order Flow

| Num | Variable              | Descriptions  | Clarification   | Format  | Example      | Example 2   |
|-----|-----------------------|---|---|---------|--------------|-------------|
| 160 | BuyNumTrades_Retail   | Total Number of Retail Buys (During Market Hours)           | Classified by the algorithm in <a href="#">Boehmer et al. (2021, JF)</a> ; See formula 35 for trade classification detail; Series available since 2006 October, See FAQ 9 and 10 before using this series in your research. | BEST32. | 9715         | 1398        |
| 161 | BuyVol_Retail         | Sum of Retail Buys, Volume in Shares (During Market Hours)  |   | BEST32. | 1886366      | 129893      |
| 162 | buy_dv_Retail         | Sum of Retail Buys, Value in Dollars (During Market Hours)  |   | BEST32. | 417980617.17 | 19401402.76 |
| 163 | SellNumTrades_Retail  | Total Number of Retail Sells (During Market Hours)          |   | BEST32. | 9434         | 1322        |
| 164 | SellVol_Retail        | Sum of Retail Sells, Volume in Shares (During Market Hours) |   | BEST32. | 1792713      | 137702      |
| 165 | sell_dv_Retail        | Sum of Retail Sells, Value in Dollars (During Market Hours) |   | BEST32. | 397177236.14 | 20555726.05 |
| 166 | total_trade_retail    | Total Number of Retail trades (During Market Hours)         |   | BEST32. | 19149        | 2720        |
| 167 | total_vol_retail      | Total Share Volume of Retail Trades (During Market Hours)   |   | BEST32. | 3679079      | 267595      |
| 168 | total_dv_retail       | Total Dollar Value of Retail Trades (During Market Hours)   |   | BEST32. | 815157853.31 | 39957128.81 |
| 169 | avg_buy_price_Retail  | Average Trade Price of Retail Buys (During Market Hours)    |   | BEST32. | 221.649      | 149.272     |
| 170 | avg_sell_price_Retail | Average Trade Price of Retail Sells (During Market Hours)   |   | BEST32. | 221.655      | 149.263     |
| 171 | bs_ratio_retail_num   | Absolute Percent Order Imbalance                            |   | BEST32. | 0.014674     | 0.027941    |

|     |                       | - Num of Retail Trades   |   |         |               |              |
|-----|-----------------------|--|---|---------|---------------|--------------|
| 172 | bs_ratio_retail_vol   | Absolute Percent Order Imbalance - Retail Trade Share Volume           |   | BEST32. | 0.025456      | 0.029182     |
| 173 | BuyNumTrades_Inst20k  | Total Number of 20K Institutional Buys (During Market Hours)           | <p>See <a href="#">Lee and Radhakrishna (2000, JFM)</a> showing 53% of institutional trades are above \$20,000 in value; <a href="#">Bhattacharya, Black, and Christensen (2007)</a> and <a href="#">Shanthikumar (2004)</a> using the \$50,000 dollar value-based cutoff.*</p> <p>*User should be aware that more recent studies (see <a href="#">Campbell et al. (2009, JFE)</a> and <a href="#">Cready et al (2014, JAR)</a>) show institutions employing algorithmic trading strategies and order-splitting techniques to disguise their trades. Therefore, orders in large sizes are merely one of the proxies for institutional trades and cannot capture all institutional trades. Its effectiveness could be challenged, given the rise of institutional algorithm trading.</p> | BEST32. | 41118         | 1385         |
| 174 | BuyVol_Inst20k        | Sum of 20K Institutional Buys, Volume in Shares (During Market Hours)  |   | BEST32. | 7788177       | 576722       |
| 175 | buy_dv_Inst20k        | Sum of 20K Institutional Buys, Value in Dollars (During Market Hours)  |   | BEST32. | 1725656760.37 | 86139059.07  |
| 176 | SellNumTrades_Inst20k | Total Number of 20K Institutional Sells (During Market Hours)          |   | BEST32. | 45007         | 1389         |
| 177 | SellVol_Inst20k       | Sum of 20K Institutional Sells, Volume in Shares (During Market Hours) |   | BEST32. | 8410432       | 540366       |
| 178 | sell_dv_Inst20k       | Sum of 20K Institutional Sells, Value in Dollars (During Market Hours) |   | BEST32. | 1863040556.88 | 80718177.84  |
| 179 | total_trade_Inst20k   | Total Number of 20K Institutional Trades (During Market Hours)         |   | BEST32. | 86125         | 2774         |
| 180 | total_vol_Inst20k     | Total Share Volume of 20K Institutional Trades (During Market Hours)   |   | BEST32. | 16198609      | 1117088      |
| 181 | total_dv_Inst20k      | Total Dollar Value of 20K Institutional Trades (During Market Hours)   |   | BEST32. | 3588697317.25 | 166857236.91 |
| 182 | avg_buy_price_Inst20k | Average Trade Price of 20K   |   | BEST32. | 221.581       | 149.270      |

|     |                        |   |  |         |              |             |
|-----|------------------------|---|--|---------|--------------|-------------|
|     |                        | Institutional Buys<br>(During Market Hours)                             |  |         |              |             |
| 183 | avg_sell_price_inst20k | Average Trade Price of 20K Institutional Sells (During Market Hours)    |  | BEST32. | 221.601      | 149.313     |
| 184 | bs_ratio_Inst20k_num   | Absolute Percent Order Imbalance - Num of 20K Institutional Trades      |  | BEST32. | 0.045155     | 0.001442    |
| 185 | bs_ratio_Inst20k_vol   | Absolute Percent Order Imbalance - 20K Institutional Trade Share Volume |  | BEST32. | 0.038414     | 0.032545    |
| 186 | BuyNumTrades_Inst50k   | Total Number of 50K Institutional Buys (During Market Hours)            |  | BEST32. | 4752         | 274         |
| 187 | BuyVol_Inst50k         | Sum of 50K Institutional Buys, Volume in Shares (During Market Hours)   |  | BEST32. | 3796317      | 336331      |
| 188 | buy_dv_Inst50k         | Sum of 50K Institutional Buys, Value in Dollars (During Market Hours)   |  | BEST32. | 841126896.59 | 50257103.61 |
| 189 | SellNumTrades_Inst50k  | Total Number of 50K Institutional Sells (During Market Hours)           |  | BEST32. | 4645         | 304         |
| 190 | SellVol_Inst50k        | Sum of 50K Institutional Sells, Volume in Shares (During Market Hours)  |  | BEST32. | 3915655      | 306946      |
| 191 | sell_dv_Inst50k        | Sum of 50K Institutional Sells, Value in Dollars (During Market Hours)  |  | BEST32. | 867004863.64 | 45867692.58 |
| 192 | total_trade_Inst50k    | Total Number of 50K Institutional                                       |  | BEST32. | 9397         | 578         |

|            |                        |   |  |         |               |             |
|------------|------------------------|---|--|---------|---------------|-------------|
|            |                        | Trades (During Market Hours)  |  |         |               |             |
| <b>193</b> | total_vol_Inst50k      | Total Share Volume of 50K Institutional Trades (During Market Hours)    |  | BEST32. | 7711972       | 643277      |
| <b>194</b> | total_dv_Inst50k       | Total Dollar Value of 50K Institutional Trades (During Market Hours)    |  | BEST32. | 1708131760.23 | 96124796.20 |
| <b>195</b> | avg_buy_price_Inst50k  | Average Trade Price of 50K Institutional Buys (During Market Hours)     |  | BEST32. | 221.576       | 149.315     |
| <b>196</b> | avg_sell_price_Inst50k | Average Trade Price of 50K Institutional Sells (During Market Hours)    |  | BEST32. | 221.522       | 149.351     |
| <b>197</b> | bs_ratio_Inst50k_num   | Absolute Percent Order Imbalance - Num of 50K Institutional Trades      |  | BEST32. | 0.011387      | 0.051903    |
| <b>198</b> | bs_ratio_Inst50k_vol   | Absolute Percent Order Imbalance - 50K Institutional Trade Share Volume |  | BEST32. | 0.015474      | 0.045680    |

# Formula Notes

## ❖ General Notation in this formula note

- $P_k$  price of trade  $k$ ,  $SHR_k$  shares of trade  $k$ ,  $Dvol_k$  Dollar Volume of trade  $k = P_k \times SHR_k$
- Trade  $k$  is a buy order  $BUY_k$ , Trade  $k$  is a sell  $SELL_k$
- $B_k$  Bidding quote,  $A_k$  Asking quote,  $M_k$  Bid-ask mid-price  $M_k = \frac{B_k + A_k}{2}$
- $D_k = \begin{cases} +1, & \text{if trade } k \text{ is a buy} \\ -1, & \text{if trade } k \text{ is a sell} \end{cases}$

• **Formula 1** Lee and Ready 1991 (LR) Previous Nanosecond Algorithm: Classify trade  $k$  as a buy if price  $P_k > M_{t-1}$ , as a sell if  $P_k < M_{t-1}$  and tick test is used when  $P_k = M_{t-1}$

$$Trade\ k = \begin{cases} BUY_k, & \text{if } P_k > M_{t-1} \\ SELL_k, & \text{if } P_k < M_{t-1} \\ \text{use tick test,} & \text{if } P_k = M_{t-1} \end{cases},$$

where  $t$  denotes millisecond (before 20150725), microsecond (between 20150725 or 20150803 to 20161023), or nanosecond (for UTP after 20161023).

## • **Formula 2** Tick Test

Lee and Ready (1991): The tick test is a technique which refers the direction of a trade by comparing its price to the price of the preceding trade(s). The test classifies each trade into four categories: an uptick, a downtick, a zero-uptick, and a zero-downtick. A trade is an uptick (downtick) if the price is higher (lower) than the price of the previous trade. When the price is the same as the previous trade (a zero tick), if the last price change was an uptick, then the trade is a zero-uptick. Similarly, if the last price change was a downtick, then the trade is a zero-downtick. A trade is classified as a buy if it occurs on an uptick or a zero-uptick; otherwise it is classified as a sell.

**Note:** WRDS classification use Lee and Ready when trade price is inside prevailing NBBO and use tick test when trade price is outside prevailing NBBO

## • **Formula 3** Simple Averaged Dollar Effective Spread

$$Dollar\ Effective\ Spread_k = 2D_k(P_k - M_k).$$

The Simple Averaged Dollar Effective Spread of stock  $i$  on day  $T$  is calculated as

$$ESpreadDollar\_AVG_{i,T} = \frac{1}{N} * \sum_{k=1}^n Dollar\ Effective\ Spread_k,$$

where  $N$  is the total number of trades of stock  $i$  on day  $T$ .



- **Formula 4** Simple Averaged Percentage Effective Spread

$$\text{Percent Effective Spread}_k = \frac{2D_k(P_k - M_k)}{M_k}.$$

The Simple Averaged Percentage Effective Spread of stock  $i$  on day  $T$  is calculated as

$$ESpreadPct\_AVG_{i,T} = \frac{1}{N} * \sum_{k=1}^n \text{Percent Effective Spread}_k,$$

where  $N$  is the total number of trades of stock  $i$  on day  $T$ .

- **Formula 5** Simple Averaged Dollar Realized Spread

$$\text{Dollar Realized Spread}_k = 2D_k(P_k - M_{k+5}),$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$RSpreadDollar\_AVG_{i,T} = \frac{1}{N} * \sum_{k=1}^n \text{Dollar Realized Spread}_k,$$

where  $N$  is the total number of trades of stock  $i$  on day  $T$ .

- **Formula 6** Simple Averaged Percentage Realized Spread

$$\text{Percent Realized Spread}_k = \frac{2D_k(P_k - M_{k+5})}{M_k},$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$RSpreadPct\_AVG_{i,T} = \frac{1}{N} * \sum_{k=1}^n \text{Percent Realized Spread}_k,$$

where  $N$  is the total number of trades of stock  $i$  on day  $T$ .

- **Formula 7** Simple Averaged Dollar Price Impact

$$\text{Dollar Price Impact}_k = 2D_k(M_{k+5} - M_k),$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$\text{PriceImpactDollar\_AVG}_{i,T} = \frac{1}{N} * \sum_{k=1}^n \text{Dollar Price Impact}_k,$$

where  $N$  is the total number of trades of stock  $i$  on day  $T$ .

- **Formula 8** Simple Averaged Percentage Price Impact

$$\text{Percent Price Impact}_k = \frac{2D_k(M_{k+5} - M_k)}{M_k},$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$\text{PriceImpactPct\_AVG}_{i,T} = \frac{1}{N} * \sum_{k=1}^n \text{Percent Price Impact}_k,$$

where  $N$  is the total number of trades of stock  $i$  on day  $T$ .

- **Formula 9** Value Weighted Average Price of Buy Trades

$$\text{BuyVWAP}_{i,T} = \sum_{k=1}^n w_k * P_k \mid k \in \{\text{BUY}_{i,T}\},$$

where  $w_k = \frac{P_k \times \text{SHR}_k}{\sum_{k=1}^n P_k \times \text{SHR}_k}$ .

- **Formula 10** Value Weighted Average Price of Sell Trades

$$\text{SellVWAP}_{i,T} = \sum_{k=1}^n w_k * P_k \mid k \in \{\text{SELL}_{i,T}\},$$

where  $w_k = \frac{P_k \times \text{SHR}_k}{\sum_{k=1}^n P_k \times \text{SHR}_k}$ .

- **Formula 11** Value Weighted Dollar Effective Spread

$$\text{Dollar Effective Spread}_k = 2D_k(P_k - M_k)$$

The Value Averaged Dollar Effective Spread of stock  $i$  on day  $T$  is calculated as

$$\text{ESpreadDollar\_VW}_{i,T} = \sum_{k=1}^n w_k \times \text{Dollar Effective Spread}_k,$$

where  $w_k = \frac{P_k \times \text{SHR}_k}{\sum_{k=1}^n P_k \times \text{SHR}_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 12** Share Weighted Dollar Effective Spread

$$\text{Dollar Effective Spread}_k = 2D_k(P_k - M_k)$$

The Share Averaged Dollar Effective Spread of stock  $i$  on day  $T$  is calculated as

$$ESpreadDollar\_SW_{i,T} = \sum_{k=1}^n w_k \times \text{Dollar Effective Spread}_k,$$

where  $w_k = \frac{SHR_k}{\sum_{k=1}^n SHR_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 13** Value Weighted Percentage Effective Spread

$$\text{Percent Effective Spread}_k = \frac{2D_k(P_k - M_k)}{M_k}$$

The Value Averaged Percentage Effective Spread of stock  $i$  on day  $T$  is calculated as

$$ESpreadPct\_VW_{i,T} = \sum_{k=1}^n w_k \times \text{Percent Effective Spread}_k,$$

where  $w_k = \frac{P_k \times SHR_k}{\sum_{k=1}^n P_k \times SHR_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 14** Share Weighted Percentage Effective Spread

$$\text{Percent Effective Spread}_k = \frac{2D_k(P_k - M_k)}{M_k}$$

The Share Averaged Percentage Effective Spread of stock  $i$  on day  $T$  is calculated as

$$ESpreadPct\_SW_{i,T} = \sum_{k=1}^n w_k \times \text{Percent Effective Spread}_k,$$

where  $w_k = \frac{SHR_k}{\sum_{k=1}^n SHR_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 15** Share Weighted Dollar Realized Spread

$$\text{Dollar Realized Spread}_k = 2D_k(P_k - M_{k+5}),$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$RSpreadDollar\_SW_{i,T} = \sum_{k=1}^n w_k \times \text{Dollar Realized Spread}_k,$$

where  $w_k = \frac{SHR_k}{\sum_{k=1}^n SHR_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 16** Value Weighted Dollar Realized Spread

$$\text{Dollar Realized Spread}_k = 2D_k(P_k - M_{k+5}),$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$R\text{SpreadDollar\_VW}_{i,T} = \sum_{k=1}^n w_k \times \text{Dollar Realized Spread}_k,$$

where  $w_k = \frac{P_k \times \text{SHR}_k}{\sum_{k=1}^n P_k \times \text{SHR}_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 17** Share Weighted Percentage Realized Spread

$$\text{Percent Realized Spread}_k = \frac{2D_k(P_k - M_{k+5})}{M_k},$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$R\text{SpreadPct\_SW}_{i,T} = \sum_{k=1}^n w_k \times \text{Percent Realized Spread}_k,$$

where  $w_k = \frac{\text{SHR}_k}{\sum_{k=1}^n \text{SHR}_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 18** Value Weighted Percentage Realized Spread

$$\text{Percent Realized Spread}_k = \frac{2D_k(P_k - M_{k+5})}{M_k},$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$R\text{SpreadPct\_VW}_{i,T} = \sum_{k=1}^n w_k \times \text{Percent Realized Spread}_k,$$

where  $w_k = \frac{P_k \times \text{SHR}_k}{\sum_{k=1}^n P_k \times \text{SHR}_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 19** Value Weighted Dollar Price Impact

$$\text{Dollar Price Impact}_k = 2D_k(M_{k+5} - M_k),$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$\text{PriceImpactDollar\_VW}_{i,T} = \sum_{k=1}^n w_k \times \text{Dollar Price Impact}_k,$$

where  $w_k = \frac{P_k \times \text{SHR}_k}{\sum_{k=1}^n P_k \times \text{SHR}_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 20** Share Weighted Dollar Price Impact

$$\text{Dollar Price Impact}_k = 2D_k(M_{k+5} - M_k),$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$\text{PriceImpactDollar\_SW}_{i,T} = \sum_{k=1}^n w_k \times \text{Dollar Price Impact}_k,$$

where  $w_k = \frac{SHR_k}{\sum_{k=1}^n SHR_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 21** Value Weighted Percentage Price Impact

$$\text{Percent Price Impact}_k = \frac{2D_k(M_{k+5} - M_k)}{M_k},$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$\text{PriceImpactPct\_VW}_{i,T} = \sum_{k=1}^n w_k \times \text{Percent Price Impact}_k,$$

where  $w_k = \frac{P_k \times SHR_k}{\sum_{k=1}^n P_k \times SHR_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 22** Share Weighted Percentage Price Impact

$$\text{Percent Price Impact}_k = \frac{2D_k(M_{k+5} - M_k)}{M_k},$$

where  $M_{k+5}$  is the bid-ask mid-point five minutes after the  $k$ th trade.

$$\text{PriceImpactPct\_SW}_{i,T} = \sum_{k=1}^n w_k \times \text{Percent Price Impact}_k,$$

where  $w_k = \frac{SHR_k}{\sum_{k=1}^n SHR_k}$ ,  $n$  is all the trades of stock  $i$  on day  $T$ .

- **Formula 23** Trade-based Intraday Volatility

The trade-based intraday volatility of stock  $i$  on day  $T$  is calculated as:

$$IVol_{i,T} = \frac{\sum_{t=1}^T (Ret_{i,t} - \overline{Ret}_{i,t})^2}{T - 1},$$

where  $Ret_{i,t} = \ln \frac{P_{k,t}}{P_{k-1,t-1}}$ , and  $P_{k,t}$  is the trade price of trade  $k$ .

Note that the “Second Intraday Indicators” use the end-of-second rule and takes the last trade of the stock in a given second when computing second-to-second returns. Multiple trades in the same second will be compressed to using the last trade of the second. Seconds with no trades are not populated.

Millisecond Intraday Indicators use neither end-of second rule nor end-of-millisecond rule. Millisecond IID treats each trade observation as an event, and therefore, the trade-based volatility measure is event-based and does not populate for each second/millisecond.

- **Formula 24** NBBO Quote-based Intraday Volatility

The quote-based intraday volatility of stock  $i$  on day  $T$  is calculated as:

$$IVol_{i,T} = \frac{\sum_{t=1}^T (Ret_{i,t} - \overline{Ret}_{i,t})^2}{T - 1},$$

where  $Ret_{i,t} = \ln \frac{M_{i,t}}{M_{i,t-1}}$ ,  $M_{i,t} = \frac{B_{i,t} + A_{i,t}}{2}$  is the bid-ask mid-price for stock  $i$  at second  $t$ .

Note that the “Second Intraday Indicators” use the end-of-second rule and takes the last NBBO update of a stock in a given second when computing second-to-second returns. Multiple NBBO updates in the same second will be compressed to using the last NBBO of the second. Seconds with no NBBO updates are not populated.

Millisecond Intraday Indicators use neither end-of second rule nor end-of-millisecond rule. Millisecond IID treats each NBBO observation as a quote innovation event, and therefore, the volatility measure is event-based and does not populate for each second/millisecond.

- **Formula 25** Variance Ratio 1 (15-second/3\*5-second)

$$VarianceRatio\ 1 = \left| \frac{VAR(Ret_{15t})}{3 \times VAR(Ret_{5t})} - 1 \right|,$$

where  $VAR(Ret_{15t})$  is the variance of 15 second log returns.

- **Formula 26** Variance Ratio 2 (1-min/4\*15-second)

$$VarianceRatio\ 2 = \left| \frac{VAR(Ret_{60t})}{4 \times VAR(Ret_{15t})} - 1 \right|,$$

where  $VAR(Ret_{60t})$  is the variance of 1-minute log returns.

- **Formula 27** Variance Ratio 3 (5-min/5\*1-min)

$$VarianceRatio\ 3 = \left| \frac{VAR(Ret_{300t})}{5 \times VAR(Ret_{60t})} - 1 \right|,$$

where  $VAR(Ret_{300t})$  is the variance of 5-minute log returns.

- **Formula 28** Variance Ratio 4 (15-min/3\*5-min)

$$\text{VarianceRatio 4} = \left| \frac{\text{VAR}(\text{Ret}_{900t})}{3 \times \text{VAR}(\text{Ret}_{300t})} - 1 \right|,$$

where  $\text{VAR}(\text{Ret}_{900t})$  is the variance of 15-minute log returns.

- **Formula 29** Variance Ratio 5 (30-min/2\*15-min)

$$\text{VarianceRatio 5} = \left| \frac{\text{VAR}(\text{Ret}_{1800t})}{2 \times \text{VAR}(\text{Ret}_{900t})} - 1 \right|,$$

where  $\text{VAR}(\text{Ret}_{1800t})$  is the variance of 30-minute log returns.

Note that the Variance Ratios in the Millisecond Intraday Indicators will receive an update in 2020 May to be comparable to the variance ratios in the Second Intraday Indicators Dataset.

- **Formula 30** Lambda (Price Impact Coefficient), Intercept Suppressed  
Lambda ( $\lambda$ ) is the regression coefficient of the following model:

$$\text{Ln} \frac{M_{i,t}}{M_{i,t-300}} = \lambda * \text{SSqrtDVol} + \epsilon,$$

where  $\text{SSqrtDVol} = \text{Sgn}(\sum_{t-300}^t \text{BuyDollar} - \sum_{t-300}^t \text{SellDollar}) \times \sqrt{|\sum_{t-300}^t \text{BuyDollar} - \sum_{t-300}^t \text{SellDollar}|}$ ,  $M_{i,t} = \frac{B_{i,t} + A_{i,t}}{2}$  is the bid-ask mid-price for stock  $i$  at second  $t$ .

- **Formula 31** Lambda (Price Impact Coefficient), Intercept Not Suppressed  
Lambda ( $\lambda$ ) is the regression coefficient of the following model:

$$\text{Ln} \frac{M_{i,t}}{M_{i,t-300}} = \alpha + \lambda * \text{SSqrtDVol} + \epsilon,$$

where  $\text{SSqrtDVol} = \text{Sgn}(\sum_{t-300}^t \text{BuyDollar} - \sum_{t-300}^t \text{SellDollar}) \times \sqrt{|\sum_{t-300}^t \text{BuyDollar} - \sum_{t-300}^t \text{SellDollar}|}$ ,  $M_{i,t} = \frac{B_{i,t} + A_{i,t}}{2}$  is the bid-ask mid-price for stock  $i$  at second  $t$ .

- **Formula 32** Herfindahl Index

$$\text{HIndex} = \frac{\sum_{t=1}^{t=1800} \sum_{k=1}^N (P_k \times \text{SHR}_k)^2}{(\sum_{t=1}^{t=1800} \sum_{k=1}^N P_k \times \text{SHR}_k)^2}$$

- **Formula 33** Absolute Percent Order Imbalance (N. of Trades)

$$OI = \frac{|\text{Total Number of Buys} - \text{Total Number of Sells}|}{\text{Total Number of Trades}}$$

- **Formula 34** Absolute Percent Order Imbalance (Volume)

$$OI = \frac{|\text{Total Vol of Buys} - \text{Total Vol of Sells}|}{\text{Total Trade Volume}}$$

- **Formula 35** Retail Trade Classification

(Applying to trades reported by FINRA only, exchange ID= "D")

Following Boehmer et al. (2021, JF), the algorithm classifies trades reporting to FINRA with fraction of penny price improvements. See an example with AMC trades below:

| Date of quote | Trade Time in milliseconds (microsecs starting 7/25/2015, nanosecs starting 10/24/2016) | Security symbol root | Security symbol suffix | type | Exchange that issued the trade | Trade Sale Condition (up to 4 codes) | Volume of trade | Price of trade |
|---------------|---|----------------------|------------------------|------|--------------------------------|--------------------------------------|-----------------|----------------|
| 20210602      | 10:33:42.913158912  | AMC                  |                        | T    | D                              | I                                    | 14              | 40.5299        |
| 20210602      | 10:33:42.921133568  | AMC                  |                        | T    | D                              | I                                    | 5               | 40.5299        |
| 20210602      | 10:33:42.928395008  | AMC                  |                        | T    | D                              | I                                    | 3               | 40.5299        |
| 20210602      | 10:33:42.955975680  | AMC                  |                        | T    | D                              | I                                    | 15              | 40.5299        |
| 20210602      | 10:33:42.965568256  | AMC                  |                        | T    | D                              | I                                    | 6               | 40.5001        |
| 20210602      | 10:33:43.039198208  | AMC                  |                        | T    | D                              | I                                    | 1               | 40.5195        |
| 20210602      | 10:33:43.040761856  | AMC                  |                        | T    | D                              | I                                    | 8               | 40.5106        |
| 20210602      | 10:33:43.087192064  | AMC                  |                        | T    | D                              | I                                    | 10              | 40.5195        |
| 20210602      | 10:33:43.138007808  | AMC                  |                        | T    | D                              | I                                    | 1               | 40.4901        |
| 20210602      | 10:33:43.155794944  | AMC                  |                        | T    | D                              | I                                    | 2               | 40.5299        |
| 20210602      | 10:33:43.178614784  | AMC                  |                        | T    | D                              | I                                    | 9               | 40.4901        |
| 20210602      | 10:33:43.206710784  | AMC                  |                        | T    | D                              |                                      | 100             | 40.528         |
| 20210602      | 10:33:43.224252416  | AMC                  |                        | T    | D                              | I                                    | 1               | 40.5299        |

Let  $Penny_k = 100 * \text{mod}(\text{price}, 0.01)$ , then

$$\text{Trade } k = \begin{cases} \text{BUY}_k, & \text{if } 1 > Penny_k > 0.6 \\ \text{SELL}_k, & \text{if } 0 < Penny_k < 0.4 \\ \text{dropped}, & \text{else} \end{cases}$$



# Frequently Asked Questions (FAQs)

1. What is the difference of Second-timestamped IID and Millisecond-timestamped IID?

The Second-timestamped IID is developed based on the NYSE Monthly TAQ data, which has second-to-second timestamps. The Millisecond-timestamped IID is built on the NYSE Daily TAQ data, which has the millisecond-level timestamps. These two datasets vary in the following perspectives:

- (1) Data range: monthly TAQ goes from 1993 to 2014 and is discontinued after 2014; daily TAQ covers from 2003 to present
- (2) Data timestamp: Monthly TAQ is timestamped at second level while Daily TAQ is at millisecond level. WRDS noticed that Monthly TAQ simply floor the millisecond timestamp to the second level, which is not a rounding method. For example,
  - Daily TAQ time: 16:00:00:500000001
  - Monthly TAQ time (same observation, flooring to the second): 16:00:00 (flooring is like truncating the finer levels)
  - Rounding method would be: 16:00:01 (this is just for comparison, it is not used in monthly TAQ)
- (3) Official NBBO: Monthly TAQ does not come with any official NBBO files from NYSE. All the NBBO and WCT files are home-made. NBBOs are generated by comparing the best bid/offer prices across exchanges at the same second. Unlike Monthly TAQ, Daily TAQ comes with NYSE official NBBO files that tell you the official NBBO prices at a given millisecond during market hours. This NBBO files also helps to eliminate quotes that are not qualified for NBB or NBO according to NYSE's definition. With the official NBBO files in Daily TAQ, we expect that millisecond IID has higher-quality variables.
- (4) Additional variables and conditions: we find the trade and quote conditions in Daily TAQ are richer and better updated compared to those in the Monthly TAQ.

2. What is the difference between Second-timestamped IID version 2.0 (2019 April release) and the vintage Second-timestamped IID?

Our first version of Second-timestamped IID Beta was released in 2017 (referred as Second IID V1.0). In 2019 April, WRDS will release V2.0 Second-timestamped IID, which has the following updates:

- (1) V 2.0 has 20 more variables, for example, more variance ratio variation variables and more out-of-market-hour variables
- (2) Fixed few missing tickers reported in the vintage version

3. What methodology/academic papers do you follow to compute the variables in WRDS IID?

Our methodology and cleaning filters used to clean the TAQ data are similar to those used by Holden and Jacobsen (JF, 2014).<sup>1</sup> You may find liquidity measure values are similar to those computed by Holden and Jacobsen's SAS code.<sup>2</sup>

4. Can I grab the same variables from Second IID and Millisecond IID together?

We understand that there is a need for a long historical sample for time-series studies. Since Monthly TAQ and Daily TAQ differs in many ways mentioned in FAQ 1, we do not recommend researchers to combine variables from Millisecond IID to those from Second IID. One biggest difference would be Millisecond IID is built on Daily TAQ which contains official NBBO from NYSE, while Second IID is built on Monthly TAQ which does not have official NBBO files. The NBBO file in Second IID is home-made and is subject to inclusion of noisy messages.

5. Why there are fewer variables in Millisecond IID than those in Second IID? Millisecond IID offers 198 variables while Second IID offers 289 variables. The reduction of variables is caused by better and more accurate methodology. Point A to C below are the variable categories that have the most variable reduction. Motivations are explained.

A. In the Monthly TAQ, users don't have trade sequence number and quote sequence number. Therefore, researchers have to assume a quote/trade distribution within the second and assign the trades with quotes. In the Millisecond IID (using Daily TAQ), we have the sequence numbers and no longer use these methods anymore.

This is the reason why we don't need other quote-trade matching methods like those used in the Second IID:

- Second IID variables (each method has 30 variables in Second IID):
  - 1) Previous second method
  - 2) Same second method
  - 3) Interpolated second method
- Millisecond IID variables:
  - 1) Previous millisecond method

This difference eliminates the two methods that are no longer needed in Daily TAQ, that's  $30 \times 2 = 60$  variables less.

<sup>1</sup> Holden, Craig W., and Stacey Jacobsen. "Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions." *The Journal of Finance* 69.4 (2014): 1747-1785.

<sup>2</sup> Prof. Holden shares his SAS code at: <https://kelley.iu.edu/cholden/>. Please note their code does not deal with the timestamp update from millisecond to microsecond and microsecond to nanosecond which affect the TAQ data after July 27, 2015.

The 30 variables are specified on page 12 in the manual, variable # 62 - # 92.

- B. WRDS Second IID has two set of timers (9:30 am to 16:00 pm and market open to market close). This might be confusing because market open is usually 9:30 and market close is 16:00 pm. Sometimes the open trade and close trade are 1-2 seconds ahead or after the time mark. In Millisecond IID, we use an improved algorithm: use market open time and market close time when they are not missing; if they are missing then use 9:30 am to 16:00 pm. This saves about 30 variables.
- C. As Daily TAQ does not have aftermarket hour NBBOs, Millisecond IID does not offer trade classification summaries for aftermarket hours. In monthly TAQ, there is no NBBO file. Second IID home-made all the NBBOs (during market hours and after-market hours). In daily TAQ, NYSE provides official NBBO files during market hours. Generally, there is no NBBO aftermarket hours (as this message is excluded in the NBBO file). Thus, we cannot assign a trade as a buy or as a sell after-market hours (WRDS has a plan to develop aftermarket hour home-made NBBO files). This cause Millisecond IID have 20 fewer variables.

From point A to C listed above, Millisecond IID has much fewer variables compared to those offered in Second IID. The reduction in the number of variables is caused by the access to finer data timestamp and message sequence availability in Daily TAQ.

- 6. Among the similar variables in both Millisecond IID and Second IID, which ones may I find significant difference? What causes such differences? Generally, there are three categories of variables that users may expect to see different values:
  - (a) Daily summary variables (pending illustration)
  - (b) Timestamp related variables (pending illustration)
  - (c) Quote related variables (pending illustration)
- 7. If I still want to grab similar variables in Second IID and Millisecond IID, what is the best way to do that?  
 We **do not** recommend combining variables from Millisecond IID (for post-2003) with the similar ones from Second IID (from 1993 to 2003). The variable may not be continuous and could have a jump in 2003's data. This would severely affect variables in the liquidity category. Some variables, however, have minor differences between the two IID products. We are preparing a mapping table for users to grab similar variables from both IID products. This mapping table will be released shortly. Please note doing so puts the researcher at risks of sample bias and continuity issues. WRDS is not responsible for and suggests against such actions.

8. What are the revisions to open and close trade? Why do I find some variable values slightly differ from those I downloaded from Version 1.0? Existing research treats all observations in the TAQ trade files are actual trades executed at the exchanges. However, we find a specific type of messages that have the same number of trade size and prices. We confirmed with NYSE and confirmed that those messages are status messages and are not real trades. Thus, we remove those messages in our computation.

Example 1: Duplicates of “Market Center Official Close (M)” that mirrors “Market Center Closing Trade (6)”. First, we want to show you examples of two trade messages having the exact same trade volume and prices back-to-back. One is with trade condition “6”, according to the manual that is the “Market Center Closing Trade”. The one following it has trade condition=”M”, noted “Market Center Official Close” in the TAQ manual. Below is an example for AAPL on 07/08/2020, the two messages both report the same trade volume of 1,995,168 shares of trades executed at 381.37 per share. Please note the two messages have two distinct trade sequence number that current research considers each distinct trade sequence number is a unique trade.

|        | Date of trade | Time to the millisecond (prior to 07/27/2015) or microsecond (starting 07/27/2015) | Exchange that issued the trade | Security symbol root | Security symbol suffix | Trade Sale Condition (up to 4 codes) | Volume of trade | Price of trade | Trade Sequence Number |
|--------|---------------|--|--------------------------------|----------------------|------------------------|--------------------------------------|-----------------|----------------|-----------------------|
| 323293 | 08JUL2020     | 15:59:59.941112342   | Q                              | AAPL                 |                        | @ I                                  | 23              | 381.38         | 4405113               |
| 323294 | 08JUL2020     | 16:00:00.000752156   | D                              | AAPL                 |                        | @                                    | 800             | 381.3323       | 4405389               |
| 323295 | 08JUL2020     | 16:00:00.111757844   | P                              | AAPL                 |                        | @ M                                  | 251             | 381.44         | 4406530               |
| 323296 | 08JUL2020     | 16:00:00.513336319   | Q                              | AAPL                 |                        | @6 X                                 | 1995168         | 381.37         | 4407089               |
| 323297 | 08JUL2020     | 16:00:00.513350040   | Q                              | AAPL                 |                        | @ M                                  | 1995168         | 381.37         | 4407090               |
| 323298 | 08JUL2020     | 16:00:00.540739376   | D                              | AAPL                 |                        | @ T                                  | 1960            | 381.37         | 4407145               |
| 323299 | 08JUL2020     | 16:00:00.540795353   | D                              | AAPL                 |                        | @ T                                  | 630             | 381.37         | 4407146               |
| 323300 | 08JUL2020     | 16:00:00.541158372   | D                              | AAPL                 |                        | @ T                                  | 554             | 381.37         | 4407147               |

Another example, IVV (an SP500 ETF) on the same day, also shows the two messages being identical.

|        | Date of trade | Time to the millisecond (prior to 07/27/2015) or microsecond (starting 07/27/2015) | Exchange that issued the trade | Security symbol root | Security symbol suffix | Trade Sale Condition (up to 4 codes) | Volume of trade | Price of trade | Trade Sequence Number |
|--------|---------------|--|--------------------------------|----------------------|------------------------|--------------------------------------|-----------------|----------------|-----------------------|
| 357422 | 08JUL2020     | 16:00:00.007187310   | P                              | IVV                  |                        | F I                                  | 37              | 317.35         | 45500002              |
| 357423 | 08JUL2020     | 16:00:00.007187310   | P                              | IVV                  |                        | F                                    | 1663            | 317.35         | 45500003              |
| 357424 | 08JUL2020     | 16:00:00.141163392   | P                              | IVV                  |                        | 6                                    | 618274          | 317.33         | 45525701              |
| 357425 | 08JUL2020     | 16:00:00.168275037   | P                              | IVV                  |                        | M                                    | 618274          | 317.33         | 45528101              |
| 357426 | 08JUL2020     | 16:00:00.173386445   | P                              | IVV                  |                        | FT                                   | 1000            | 317.34         | 45528201              |

Please note in the two examples above, the trade volume in the highlighted trade is about 6.3% of the total trade volume for Apple, and 10% for IVV on 07/08/2020.

The problem is worse after market hours for ETFs. I find the message with “Market Center Official Close (M)” trade condition is reported multiple times. All have the same trade volume and trade price, with a unique trade sequence number assigned by SIP.

|   | Date of trade | Time to the millisecond (prior to 07/27/2015) or microsecond (starting 07/27/2015) | Exchange that issued the trade | Security symbol root | Security symbol suffix | Trade Sale Condition (up to 4 codes) | Volume of trade | Price of trade | Trade Sequence Number |
|---|---------------|--|--------------------------------|----------------------|------------------------|--------------------------------------|-----------------|----------------|-----------------------|
| 1 | 08JUL2020     | 16:00:00.141163392   | P                              | IVV                  |                        | 6                                    | 618274          | 317.33         | 45525701              |
| 2 | 08JUL2020     | 16:00:00.168275037   | P                              | IVV                  |                        | M                                    | 618274          | 317.33         | 45528101              |
| 3 | 08JUL2020     | 16:10:00.001393019   | P                              | IVV                  |                        | M                                    | 618274          | 317.33         | 45587903              |
| 4 | 08JUL2020     | 18:30:00.002842062   | P                              | IVV                  |                        | M                                    | 618274          | 317.33         | 45818908              |
| 5 | 08JUL2020     | 20:00:00.001058136   | P                              | IVV                  |                        | M                                    | 618274          | 317.33         | 45895801              |

And the timestamp pattern is not unique for a given ETF, it is the same timestamp pattern for all ETFs. Note that Exchange ID = “P” is NYSE Arca, which is the major marketplace for ETF trading.

|    | Date of trade | Time to the millisecond (prior to 07/27/2015) or microsecond (starting 07/27/2015) | Exchange that issued the trade | Security symbol root | Security symbol suffix | Trade Sale Condition (up to 4 codes) | Volume of trade | Price of trade | Trade Sequence Number |
|----|---------------|--|--------------------------------|----------------------|------------------------|--------------------------------------|-----------------|----------------|-----------------------|
| 1  | 08JUL2020     | 16:00:00.168275037   | P                              | IVV                  |                        | M                                    | 618274          | 317.33         | 45528101              |
| 2  | 08JUL2020     | 16:10:00.001393019   | P                              | IVV                  |                        | M                                    | 618274          | 317.33         | 45587903              |
| 3  | 08JUL2020     | 18:30:00.002842062   | P                              | IVV                  |                        | M                                    | 618274          | 317.33         | 45818908              |
| 4  | 08JUL2020     | 20:00:00.001058136   | P                              | IVV                  |                        | M                                    | 618274          | 317.33         | 45895801              |
| 5  | 08JUL2020     | 16:00:00.222444232   | P                              | SPY                  |                        | M                                    | 804208          | 316.18         | 56534301              |
| 6  | 08JUL2020     | 16:10:00.000466871   | P                              | SPY                  |                        | M                                    | 804208          | 316.18         | 56888601              |
| 7  | 08JUL2020     | 18:30:00.000551371   | P                              | SPY                  |                        | M                                    | 804208          | 316.18         | 57538601              |
| 8  | 08JUL2020     | 20:00:00.002622517   | P                              | SPY                  |                        | M                                    | 804208          | 316.18         | 57755801              |
| 9  | 08JUL2020     | 16:00:00.148734937   | P                              | VOO                  |                        | M                                    | 27554           | 290.38         | 49884701              |
| 10 | 08JUL2020     | 16:10:00.004288467   | P                              | VOO                  |                        | M                                    | 27554           | 290.38         | 50003401              |
| 11 | 08JUL2020     | 18:30:00.003928749   | P                              | VOO                  |                        | M                                    | 27554           | 290.38         | 50187402              |
| 12 | 08JUL2020     | 20:00:00.006821314   | P                              | VOO                  |                        | M                                    | 27554           | 290.38         | 50243601              |

To adjust for those status messages, we confirm that observations in the TAQ Trades file with trade condition = M “Market Center Official Close” are not real trades. They are more sort of a status message produced by the Security Information Processor (SIP). Therefore, such messages should not be counted when computing daily summary trade volume related variables.

This modification also helps to narrow the daily share volume gap with CRSP. After removing open trade status messages “Q” and close trade status messages “M”, the difference between daily volume in IID (4am to 8pm, total\_vol\_b+ total\_vol\_m+ total\_vol\_a) and CRSP Daily share volume (VOL) is now less than 0.7%.

- Boehmer et al. (2021, JF) Stopped their sample period in 2015, is it safe to use the retail classification after this point?

Boehmer et al. (2021, JF) stopped their sample in 2015. On page 5 of their paper, they mentioned “*Second, from 2016 to September 2018, the SEC adopted a tick size pilot program (TSPP) that affects tick size and brokers’ ability to provide price improvement for many stocks, which likely affects the prevalence of subpenny price improvements unevenly in the cross section.*”



We checked the details of the TSPP and find that it affects 1,200 selected stocks from 10/01/2016 to 10/01/2018.

The details of the program can be found here:

- SEC: <https://www.sec.gov/ticksizepilot>
- FINRA: <https://www.finra.org/rules-guidance/key-topics/tick-size-pilot-program>

FINRA collected data during the program: <https://www.finra.org/rules-guidance/key-topics/tick-size-pilot-program/data-collection-securities-and-pilot-securities-files>

The percent of penny rule may not work for the stocks selected in this program.

The list of 1,200 effected stocks can be found here:

[http://tsp.finra.org/finra\\_org/ticksizepilot/TSPilotChanges.txt](http://tsp.finra.org/finra_org/ticksizepilot/TSPilotChanges.txt).

After the expiration of the program on October 2018, the limits posted by the is no longer affecting the 1,200 treatment stocks. So the classification algorithm should work for those 1,200 stocks again after 2018. (See <https://www.sec.gov/news/public-statement/tm-dera-expiration-tick-size-pilot>)

*“Therefore, on Sept. 10, 2018, the Commission issued an exemption to the exchanges and Financial Industry Regulatory Authority (FINRA) to allow the quoting and trading requirements to terminate at the end of trading on Friday, Sept. 28, 2018.”*

So far, we did not find any regulation changes related to the Tick Size Pilot Program (TSPP). We will keep an eye on it. It also means, other than the treatment 1,200 stocks that may have problem with the retail order flow identification during Oct 2016 to Oct 2018, the rest of the universe still works with the algorithm posted by Boehmer et al. (2021, JF).

10. What are the challenges I should be aware of when using the retail classification by Boehmer et al. (2021, JF)?

We previously linked an article about [Robinhood sells its marketable order flow to citadel](#) and are concerned that not all retail order flows are captured by the algorithm. We want to thank Professor Charles Jones for clarifying this: “we do not capture every retail order that is marketable, only the ones that are executed off exchange and price improved by a small fraction of a penny. Although Robinhood sells its marketable order flow to citadel and other HFT firms, that order flow is usually directly filled by citadel or the purchaser of the flow, but regardless of how it’s executed

after Robinhood sells the order, the off exchange trade must be reported to a FINRA TRF and would thus be captured by our algorithm as long as it's price improved by a small fraction of a penny." We suggest users check BJZZ<sup>3</sup> for more details on this and cite the BJZZ paper when using the retail order flows variables.

# In-development Updates

- Release Note 2019 March:

Millisecond IID Version 1.0 (computed from Daily millisecond-timestamped TAQ, 2003-2019):

- Released Complete NBBO files
- Released WCT files
- Released Daily Summaries

- Update Note 2021 July

Millisecond IID:

- Update to Version 2.0
- Revisions to the open/close trade classifications
- Removed noises from the intraday timestamp residuals
  - All revisions are highlighted in **bright red** throughout the variable lists
- Added the module of retail and institutional order flow
  - See Variable List IX, and FAQs 9 and 10

Planned Future Products

- Coming late 2021:

Millisecond IID PLUS:

- Release 5-minute/15-minute/30-minute intraday summaries
- Daily Classic PIN computed from 5-minute intervals
- WRDS-enhanced trade classification algorithm

- Coming 2022 Summer:

Millisecond IID Integrated with WRDS Event Study:

- Daily event study using WRDS IID
- Intraday event study using WRDS IID PLUS

WRDS users: We sincerely thank the feedback and suggestions from our users for proposing new variables. Due to our server capability and plans to load other high-frequency data in the future, we are not able to accommodate all requests. Some suggested variables can be computed using the intraday interval summaries that will be released soon. Thank you for your understanding. We are still open to new variable suggestions using the Daily TAQ data for empirical financial research. If you have variables that you would like to suggest, please contact the WRDS research team at [jun5@wharton.upenn.edu](mailto:jun5@wharton.upenn.edu). We will give high priority to variables that are computing intensive and demanded by large empirical research audience.