Lee and Ready (1991) Algorithm

In the market microstructure study, Lee and Ready (1991) methodology is widely used to infer trade directions: whether a trade is buyer-initiated or seller-initiated. In order to expedite trade classification algorithms, WRDS has generated two sets of data that rely on original TAQ Trades and Quotes datasets: the first set consists of the NBBOs that are derived each second from individual NBBO eligible quotes. See the NBBO Research Application (https://wrds-web.wharton.upenn.edu/wrds/research/applications/microstructure/NBBO%20derivation/) for more details on how NBBO data is generated and more details on the assumptions made. Then, the NBBOs Bid and Ask midpoints are matched to each trades at seconds 0, -1, -2 and -5 relative to trade time, and stored along with trades in the same datasets, name WRDS-derived Trades files (WCT datasets). Such approach provides users with all necessary components to infer the trade directions regardless of what specifications and assumptions are employed on the trade-quote lag, or on trade filters.

The program below relies on WRDS-derived WCT files which has matched trades and NBBO quotes. These WCT files are located under /wrds/nyse/sasdata/wrds_taqs_ct directory of the WRDS cloud server. Therefore, for the program to run, you will need to define WTAQ folder in SAS as:

```
libname WTAQ '/wrds/nyse/sasdata/wrds_taqs_ct';
```

The methodology, as originally developed by Lee and Ready (1991), involves a two-step approach: a quote test first, then a tick test. In the Quote Test, if the price of a trade is lower than the midpoint of the matched bid and ask, then the trade is classified as seller-initiated. If the price is higher, the trade is considered to be buyer-initiated. When they are equal, the tick test is utilized. In the Tick Test, if the price is lower than the previous price, it is classified as seller-initiated and if it is higher, then the trade is considered to be buyer-initiated. If the tick test results are not conclusive, then the current price is again compared with the price with two lags. Subsequently, any inclusive trades will be classified as undetermined.

Originally, Lee and Ready (1991) applies a so-called 5-second rule to adjust potential

delay of the reporting time for trades. However, the latency in quote issuance and trade execution has been decreasing over time, and more recent research has recommended the use of 2 seconds or 0 seconds lags instead. For example, Peterson and Sirri (2003), and Bessembinder (2003) recommend a 0-second delay for the reporting time for trades. In other words, they suggest don't do any adjustment for the trade's report time. On the other hand, Henker and Wang (2006) show that a 1-second rule is more appropriate.

This program is efficient in the sense that takes advantage of TAQ Quote and Trade datasets' sort order and employs the use of SAS 'Data Views'. The latter saves time and working space. For intermediate datasets, that can be very large when working with TAQ databases, SAS data Views stores only the instructions on how to construct the intermediate data. Avoiding unnecessary reading, transfer, and writing of intermediate datasets on temporary disk space, the SAS Views concept saves time on processing TAQ jobs and allows users to view and debug intermediate steps without the need to wait for large dataset to be processed. For further information about SAS Data Views, see a paper by James C. Stokes (http://www2.sas.com/proceedings/sugi29/067-29.pdf (http://www2.sas.com/proceedings/sugi29/067-29.pdf)) Also, the SAS documentation on Data Views has nice introduction and illustrations of this efficient programming concept (http://support.sas.com/onlinedoc/913/getDoc/en/lrcon.hlp/a001278762.htm (http://support.sas.com/onlinedoc/913/getDoc/en/lrcon.hlp/a001278762.htm)).

While the program does not specify any filters on trades data, it is important to note that previous research suggested several filters to subset TAQ Trades data for 'standard observations' (See Engle and Sun (2007), Jiang and Oomen (2008),Liu and Maheu (2008)). Such filters include restricting the observations to trades with positive price, uncorrected trades (correction indicator corr = 0), using the trade condition to keep only regular trades (e.g. dropping bunched trades, cond not in ("O" "Z" "B" "T" "L" "G" "W" "J" "K").

?

```
/*
*************************************************************************
*/
/* ************* W R D S   R E S E A R C H   A P P L I C A T I O N S
************* */
/*
*************************************************************************
*/
/* Summary   : Identify Buyer/Seller Initiated Trades using Lee and Ready
(1991)     */
/* Date      : May 2009, Revised September,
2011                                    */
/* Author    :
WRDS                                                                  */
/* Notes     : Lee and Ready (1991) Algorithm uses Quote Test first then Tick
Test   */
/*             Use WCT data that has matched trades and NBBO
quotes                  */
/*             WCT matched NBBO quotes are at the end of seconds 5, 2, 1, and
0      */
/*                 relative to the trade
time                                        */
/*             The code also generates Total Daily Buy and Sell Volume (&
Imbalance) */
/*
*************************************************************************
*/


/* Decide on Quote Lag: Lee and Ready (1991) use 5-Second Rule */
%let SEC=1;    /* Use Quote at the end of previous second, instead of 5-Second
Rule */

/* Select Period */
%let yymmdd=20090120;

/* Step 1. Derive Trade Sign using Lee and Ready Algorithm */
data Lee_Ready;
set WTAQ.WCT_&yymmdd;
where time between "09:30:00"t and "16:00:00"t and price>0 and size>0;
/* Lee and Ready Test */
/* Apply Quote Test first */
  LeeReady=sign(Price-MidPoint&sec);
/* Then, Apply Tick Test  */
if LeeReady=0 then LeeReady=TICK;
keep symbol date time price size LeeReady;
run;

/* Step 2. Calculate Daily Buy and Sell Volumes */
data Buy_Sell_Imbalance;
set Lee_Ready;
by symbol date time;
retain N_TRADES SIZ_SUM VAL_SUM BUYS SELLS 0;
if first.date then do; N_TRADES=0; SIZ_SUM=0; VAL_SUM=0; Buys=0; Sells=0; end;
N_TRADES+1;
SIZ_SUM+size;
VAL_SUM+size*price;
Buys+size*(LeeReady=1);
Sells+size*(LeeReady=-1);
VWAP=val_sum/siz_sum;
if last.date;
label Buys = "Total Buy Volume" Sells="Total Sell Volume" N_Trades="Number of
Daily Trades";
```

```
label SIZ_SUM="Total Daily Volume" VAL_SUM="Total Daily Dollar Volume";
label Price="Price at Day End" VWAP="Volume Weighted Average Price";
format Buys Sells SIZ_SUM N_TRADES comma12. VAL_SUM dollar12. VWAP Price
dollar8.2;
keep symbol date time buys sells VAL_SUM SIZ_SUM N_TRADES VWAP Price;
run;

/* End */

/*
**************************************************************************
*/
/* *************  Material Copyright Wharton Research Data Services
*************** */
/* ****************************** All Rights Reserved
****************************** */
/*
**************************************************************************
*/
```

It is crucial to perform the data merging and computational analysis at the smallest unit using daily datasets, and then do the looping to append the resulting output datasets. This speeds up the processing time substantially and saves disk space and processing resourses.Try as much as possible to minimize the practice of appending the input trades and quotes datasets before the analysis stage as this is very costly in terms of time, speed, and computing resources. For this reason, it is a good idea to include the date in the output name, so you can use it in the loop to append your resulting output over your sample period.

References

Bessembinder, Hendrik, 2003, "Issues in Assessing Trade Execution Costs," Journal of Financial Economics, 6, 233-257.

Engle, Zheng, and Robert F. Sun, 2007, "When is Noise Not Noise - A Microstructure Estimate of Realized Volatility," New York University Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1293630 (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1293630).

Jiang, George J., and Roel C. Oomen, 2008, "Testing for Jumps when Asset Prices are Observed with Noise–a Swap Variance Approach", Journal of Econometrics, Vol. 144 No. 2, 352-370.

Lee,Charles M. and Mark J. Ready, 1991,“Inferring Trade Direction from Intraday Data,” Journal of Finance, Vol. 46 No. 2, 733-746.

Liu, Chun, and John M. Maheu, 2008, "Are There Structural Breaks in Realized Volatility?" Journal of Financial Econometrics, Vol. 6 No, 3, 326-360.

(http://www.wharton.upenn.edu)

About WRDS (https://wrds-web.wharton.upenn.edu/wrds/about/index.cfm)
WRDS FAQs (https://wrds-web.wharton.upenn.edu/wrds/about/WRDS FAQs.cfm)
WRDS News (https://wrds-web.wharton.upenn.edu/wrds/news/index.cfm)

3 Ways to use WRDS (https://wrds-web.wharton.upenn.edu/wrds/about/three_ways_to_use_WRDS.cfm)
Account Types on WRDS (https://wrds-web.wharton.upenn.edu/wrds/about/Account Types.cfm)
Terms of Use (https://wrds-web.wharton.upenn.edu/wrds/about/terms.cfm)

Account Preferences (https://wrds-web.wharton.upenn.edu/wrds/mywrds/preferences.cfm)
Info / Support Request (https://wrds-web.wharton.upenn.edu/wrds/about/external_support_request.cfm)
Privacy Policy (https://wrds-web.wharton.upenn.edu/wrds/about/privacy.cfm)

WRDS Demo (https://wrds-web.wharton.upenn.edu/wrds/demo/)
Conference Calendar (https://wrds-web.wharton.upenn.edu/wrds/about/Conference Calendar.cfm)
Best Paper Awards (http://www.whartonwrds.com/best-paper-award-winners/)

Wharton Research Data Services