WRDS E-Learning Session
Lee and Ready Method (1991)

Dr. Yuxing Yan
Research Service Director
(3/25/2010)

Agenda

- Lee and Ready (1991, JF)
- Several questions
- Why Lee and Ready test is important?
- 5-second, 1-second or 0-second
- pre-process data
- collapse quotes within the same second
- Weighted prices when the trades happen at the same time
- program (lee_ready.sas)

What to measure?

- We observe a trade.

- `SYMBOL   DATE     TIME     PRICE  SIZE G127 CORR COND   EX`
- `IBM    19990104 9:45:09 185.000 500  0      0           N`

- Who initiated this trade?

Quote test first

- <mark>Quote test:</mark>

- Seller-initiated if price < (bid+ask)/2

- Buyer-initiated if price > (bid+ask)/2

- Note: Bid and quote are from quote entry immediate before a trade in terms of time

- Just opposite of: buyer-low and sell high

- Mid-point=(bid+ask)/2

Put trades and quote together
(is this correct?)

From trade:

SYMBOL	DATE	TIME	PRICE
IBM	19990104	9:45:09	185.000

From Quote

Obs	SYMBOL	DATE	TIME	BID	OFR	mid_point
1	IBM	19990104	9:45:02	184.688	185.125	184.906
2	IBM	19990104	9:45:06	184.500	185.375	184.938
3	IBM	19990104	9:45:06	184.750	185.000	184.875
4	IBM	19990104	9:45:08	184.625	185.125	184.875
5	IBM	19990104	9:45:09	184.500	185.375	184.938
6	IBM	19990104	9:45:09	184.500	185.375	184.938

But if apply 5-second rule
(potential delay of reporting of a trade)

From trade:

	SYMBOL	DATE	TIME	PRICE
old	IBM	19990104	9:45:09	185.000
new	IBM	19990104	9:45:04	185.000

From Quote

Obs	SYMBOL	DATE	TIME	BID	OFR	mid_point
1	IBM	19990104	9:45:02	184.688	185.125	184.906
2	IBM	19990104	9:45:06	184.500	185.375	184.938
3	IBM	19990104	9:45:06	184.750	185.000	184.875
4	IBM	19990104	9:45:08	184.625	185.125	184.875
5	IBM	19990104	9:45:09	184.500	185.375	184.938

Buyer-initiated trade

From trade

IBM   19990104   9:45:04   185.000

From quote (immediate quote before it)

IBM   19990104   9:45:02   184.906

Since 185> 184.906, buyer-initiated trade

Tick test if price=mid-point

Seller-initiated: P(t) <P(t-1)

Buyer-initiated : p(t)>P(t-1)

Where   p(t)   is the price at time t

p(t-1) is the price at time t-1, i.e.,
previous trading price

Note: users could go back many lags

wrds Lee and Ready test (1991)

Identify who initiates a trade

Quote test

Seller initiated: price < (bid + ask)/2

Buyer-initiated: price > (bid + ask)/2

When prices =(bid+ask)/2 then use tick test

Tick test

Seller-initiated: P(t) <P(t-1)

Buyer-initiated : p(t)>P(t-1)

Several questions

1) Why Lee and Ready test is important?
2) How to align trades and quotes?
3) How to treat multiple quotes within one second?
4) How to treat multiple trades happen at the same time?
5) How to process trade data?
6) How to process quote data?

Some research topics

wrds

- Adverse selection

- Market efficiency

- Transaction cost

- Liquidity

- Insider information

- Institutional trading

- Trading strategy

- Fraud detection

  Note: see my reference list related to microstructure study

One example: PIN model

- Easley et al. (1996) JF

- Input B (# of buyer-initiated trades, S; # of seller-initiated trades)
- output: α, δ, μ, ε_B, and ε_S
- α: prob. of events
- δ : prob. a bad event when an event happens
- μ : arrival rate of uninformed trade
- ε_B : arrival rate of buyer-initiated trades
- ε_S : arrival rate of seller-initiated trades

Major assumptions and logic

- Poisson process for the information arrival

$$P(n, \lambda) = \frac{e^{-\lambda} \lambda^n}{n!}$$

- Events are independent

Maximum log function

$$L(\theta \mid B_i, S_i) = \alpha(1-\delta)e^{-(\mu+\varepsilon_b)} \frac{(\mu+\varepsilon_b)^{B_i}}{B_i!} e^{-\varepsilon_s} \frac{\varepsilon_s^{S_i}}{S_i!}$$

$$+\alpha\delta e^{-\varepsilon_b} \frac{\varepsilon_b^{B_i}}{B_i!} e^{-(\mu+\varepsilon_s)} \frac{(\mu+\varepsilon_s)^{S_i}}{S_i!} + (1-\alpha)e^{-\varepsilon_b} \frac{\varepsilon_b^{B_i}}{B_i!} e^{-\varepsilon_s} \frac{\varepsilon_s^{S_i}}{S_i!}$$

PIN (prob. Information based trade)

$$L(\theta \mid M) = \prod_{i=1}^{I} L(\theta \mid B_i, S_i)$$

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_B + \varepsilon_S}$$

Notations for B and S

- B: daily number of buyer-initiated trades
- S: daily number of seller-initiated trades

- Easley et al. (1996b, 1997,1998, 2004, 2005), Brown et al. (2004), Vega (2006),Yan and Zhang (2006), Benos and Jochec (2007), Fuller et al. (2007), Aslan et al. (2007), Hameed et al. (2008) and Duarte and Young (2009) and Yan(2010)

Matching trades with quotes

- Matching by SYMBOL, DATE and TIME

- <u>Method 1:</u> 5-second rule, Lee and Ready (1991)
  - Delay of the report time for a trade
  - An isolated trade is a trade within a window just one trade in it.
    - Objective : identify the patterns of the delay of quotes entered the system.
    - An isolated trade is the first trade between 11:00am to 2:30pm with no other trades within 2-minute window (Lee and Ready,1991)

Matching trades with quotes – other rules

- Method 2: 0-second rule,
  - Peterson and Sirri (2003), and Bessembinder (2003)

- Method 3: 1-second rule
  - Henker and Wang (2005)

First 10 lines from Consolidated Trade

Obs	SYMBOL	DATE	TIME	PRICE	SIZE	G127	CORR	COND	EX	TSEQ
1	A	20000103	9:34:01	78.75	64700	40	0		N	807127
2	A	20000103	9:34:04	78.75	100	0	0		M	0
3	A	20000103	9:34:04	78.75	1000	0	0		M	0
4	A	20000103	9:34:04	78.75	100	0	0		M	0
5	A	20000103	9:34:04	78.75	200	0	0		M	0
6	A	20000103	9:34:04	78.75	100	0	0		M	0
7	A	20000103	9:34:04	78.75	100	0	0		M	0
8	A	20000103	9:34:04	78.75	100	0	0		M	0
9	A	20000103	9:34:04	78.75	100	0	0		M	0
10	A	20000103	9:34:04	78.75	100	0	0		M	0

Several variables for CT

SYMBOL   this variable is not a permanent stock

G127   Combination of following 3 rules

                              G rule: trading for its own account

                              127 rule: executed as a block position

                              Stopped stock indicator

                  e.g., G127=0, does not qualify as "G", Rule 12 or stopped stock trade

                              G127=40   A display book-reported trade

CORR   Correction indicator

                  e.g,   CORR=0   regular trade

COND   Condition of a trade

                  e.g., COND='A'   Cash-only basis

wrds

First 10 lines from Consolidated Quote

Obs	SYMBOL	DATE	TIME	BID	OFR	BIDSIZ	OFRSIZ	MODE	MMID	QSEQ
1	A	20000103	8:59:07	0.000	0.000	0	0	12	T	PTRS	0
2	A	20000103	8:59:07	0.000	0.000	0	0	12	T	SWST	0
3	A	20000103	8:59:07	0.000	0.000	0	0	12	T	TRIM	0
4	A	20000103	8:59:07	0.000	0.000	0	0	12	T	MADF	0
5	A	20000103	9:34:02	0.000	0.000	0	0	12	C		0
6	A	20000103	9:34:08	78.625	78.875	10	10	10	N		807129
7	A	20000103	9:34:10	78.500	79.000	1	1	12	X		0
8	A	20000103	9:34:10	77.750	79.750	1	1	12	C		0
9	A	20000103	9:34:12	78.500	79.000	1	1	12	T	MADF	0
10	A	20000103	9:34:12	78.500	79.000	1	1	12	T	CAES	0

Several variables for CQ

wrds

BID   Bid price

OFR   Offer price

BIDSIZ   Bid size   (100 share units)

OFRSIZ   Offer size (100 share units)

MODE   Quote condition

       e.g.   MODE=0   Invalid field

                  MODE=4   News dissemination
                                      (regulatory halt)

MMID   NASDAQ market maker

Preprocess data

1. Filtering out *invalid* trades

2. Filtering out *invalid* quotes

3. matching trades with quotes

Filtering out invalid trades

Keep if

1) price: price >0

 2) size :   size >0

 3) CORR: Correction indicator

                CORR = 0, 1 or 2

 4) COND :   Sale Condition

   COND not in ("O" "Z" "B" "T" "L" "G" "W" "J" "K"
)

CORR-correction indicator (CT)

Good Trades

0   Regular trade

1   Trades were later corrected

2   Symbol correction

Original Trade Records

7   Trade cancelled due to error

8   Trade cancelled

9   Trade cancelled due to symbol correction

Correction Instructions

10 Cancel record (associated with 8)

11 Error record (associated with 7)

12 Correction record (associated with 1)

Correction Indicator (CT0001)

CORR	Frequency	Cumulative Percent	Cumulative Frequency	Percent
0	64511091	99.66	64511091	99.66
1	34701	0.05	64545792	99.71
7	2117	0.00	64547909	99.71
8	74313	0.11	64622222	99.83
10	74313	0.11	64696535	99.94
11	2117	0.00	64698652	99.95
12	34701	0.05	64733353	100.00

Example code=1 and 12

```
Obs  SYMBOL      DATE    TIME      PRICE   SIZE  G127  CORR  COND  EX  TSEQ

  1    A      20000103  10:14:50  72.0000  100    0     1     Z    B    0
  2    A      20000103  10:15:48  72.0000  100    0    12          B    0

  3    A      20000103  11:19:03  69.2500  2800   0     1          B    0
  4    A      20000103  11:24:58  69.1875  2800   0    12          B    0

  5    A      20000103  11:33:14  70.7500  100    0     1          M    0
  6    A      20000103  11:33:31  70.5000  100    0    12          M    0

  7    A      20000103  12:55:12  71.8750  2000   0     1          B    0
  8    A      20000103  13:00:57  71.8125  2000   0    12          B    0

  9    A      20000103  15:31:06  71.3750  500    0     1          M    0
 10    A      20000103  15:31:30  71.5625  500    0    12          M    0
```

Add filters for CT

```
data trades;
      set taq.ct0001;
      where price>0 and size>0 and
      corr in    (0,1,2) and  cond not in
      ("O" "Z" "B" "T" "L" "G" "W"   "J" "K" );
run;
```

COND: condition of sale

wrds

- COND='O'
  - an opening trade that occurs in sequence but is reported to the tape in a later time
- COND='B'
  - Bunched trade (aggregate of two or more regular trades executed within 60 seconds with same price
- COND='G'
  - A bunched trade not reported within 90 seconds

wrds Filtering out invalid Quotes

Keep if

1) price: bid >0,   ofr >0

 2) size :   bidsiz>0 , ofrsiz>0

 3) mode: mode not in (4, 7, 9, 11, 13, 14, 15, 19, 20, 27, 28)

e.g.,

   mode=4: regulatory halt (news dissemination)

   mode=7: non-regulatory halt (order imbalance)

   mode=9: regulatory halt

Codes for filtering out invalid quotes

```
data quotes;
    set taq.cq0207;
    where bid>0 and ofr>0 and
     bidsiz>0 and ofrsiz >0 and
    mode not in
    (4, 7, 9, 11, 13, 14, 15, 19, 20, 27, 28);
run;
```

wrds

Obs	SYMBOL	DATE	TIME	PRICE	SIZE	G127	CORR	COND	EX	n
2	A	20080902	9:30:04	35.09	100	0	0	@	T	1
3	A	20080902	9:30:04	35.09	100	0	0	F	T	2
4	A	20080902	9:30:04	35.22	100	0	0	@	T	3
5	A	20080902	9:30:05	35.30	100	0	0	@	P	1
6	A	20080902	9:30:47	35.18	150	0	0	@	P	1
7	A	20080902	9:31:07	35.13	100	0	0	F	P	1
8	A	20080902	9:31:07	35.14	300	0	0	@	P	2
9	A	20080902	9:31:07	35.14	100	0	0	@	P	3
10	A	20080902	9:31:07	35.32	100	0	0	F	P	4
11	A	20080902	9:31:07	35.33	30800	0	0	@	N	5
12	A	20080902	9:31:08	35.16	100	0	0	F	T	1
13	A	20080902	9:31:08	35.17	100	0	0	@	T	2
14	A	20080902	9:31:08	35.17	100	0	0	F	T	3
15	A	20080902	9:31:08	35.17	100	0	0	F	T	4
16	A	20080902	9:31:08	35.20	100	0	0	F	T	5
17	A	20080902	9:31:08	35.20	200	0	0	F	T	6
18	A	20080902	9:31:08	35.20	100	0	0	@	T	7
19	A	20080902	9:31:08	35.15	100	0	0	F	T	8
20	A	20080902	9:31:08	35.15	100	0	0	F	T	9

wrds

Obs	SYMBOL	DATE	TIME	BID	OFR	BIDSIZ	OFRSIZ	MODE	EX	MMID	n
1	A	20080902	9:31:07	35.12	35.33	2	28	10	N		1
2	A	20080902	9:31:07	35.12	35.33	2	28	12	N		2
3	A	20080902	9:31:07	35.12	35.33	2	28	12	N		3
4	A	20080902	9:31:07	35.12	35.33	2	28	12	N		4
5	A	20080902	9:31:07	35.12	35.33	2	28	12	N		5
6	A	20080902	9:31:07	35.12	35.33	2	28	12	N		6
7	A	20080902	9:31:07	35.12	35.28	2	32	12	N		7
8	A	20080902	9:31:16	35.13	35.42	11	1	12	I		1
9	A	20080902	9:31:16	35.13	35.42	11	1	12	I		2
10	A	20080902	9:31:16	35.13	35.30	11	2	12	P		3
11	A	20080902	9:31:16	35.13	35.42	11	1	12	I		4
12	A	20080902	9:31:16	35.13	35.24	13	1	12	N		5
13	A	20080902	9:31:16	35.13	35.30	11	2	12	P		6
26	A	20080902	9:35:27	35.33	35.40	7	11	12	N		1
27	A	20080902	9:35:27	35.33	35.40	7	11	12	N		2
93	A	20080902	9:39:01	35.41	35.46	1	15	12	N		1
94	A	20080902	9:39:01	35.43	35.46	8	16	12	T		2
95	A	20080902	9:39:01	35.43	35.46	6	16	12	T		3
96	A	20080902	9:39:01	35.43	35.46	6	13	12	T		4
97	A	20080902	9:39:01	35.43	35.46	5	13	12	T		5

Parts of codes for Lee &Ready test

```
***************************************************************
* Step 1: Input area
***************************************************************;

%let five_second_rule=5;    * 5-second rule;
%let cttime1="9:45:00"t;    * first trade time;
%let cqtime1="9:30:00"t;    * first quote time;
%let time2="16:00:00"t;     * ending time;
%let vars_in_cq=symbol date time ofr bid mode bidsiz ofrsiz; * vars from CQ;
%let vars_in_ct=symbol date time price size cond corr;  * vars from CT;
%let stocks='IBM' 'DELL';

* Many lines here  ******;

***************************************************************
     * Step 2: Get trade data with filters
 ***************************************************************;
    data trades(drop=corr cond)/view=trades;
       set taq.ct&year&prefix.&i(keep=&vars_in_ct);
       where symbol in (&stocks) and time between &cttime1 and &time2 and
       price>0 and corr=0 and
       cond not in ( "O" "Z" "B" "T" "L" "G" "W" "J" "K" );
    run;
```

Weighted trades happen at the same time

```
*********************************************************
* Step 3: Get weighted prices for trades happened at the same time
**************************************************************,
proc means data=trades noprint;
        var price/weight=size;
        by symbol date time;
        output out=trades2(rename=(_freq_=n_trades) drop=_type_) mean=price;
 run;
proc datasets library=work; delete trades; run;
```

Process quote data

```
**********************************************************
* Step 5: Get quote changes
************************************************************;
data quotes(drop=oldmp mode bidsiz ofrsiz)/view=quotes;
    set taq.cq&year&prefix.&i(keep=&vars_in_cq);
    by symbol;
    where symbol in (&stocks) and
    nmiss(bid,ofr,bidsiz,ofrsiz)=0 and ofr>bid and
    mode not in (0,4, 7, 9, 11, 13, 14, 15, 19, 20, 27, 28);
    midpoint = (bid+ofr)/2;
    oldmp = lag(midpoint);
    if first.symbol then oldmp = .;
    label midpoint='bid-ask midpoint';
    if midpoint ne oldmp then output;
  run;
```

wrds

Lee_ready.sas

For a complete codes, please see

→ Support→ Research Application

http://wrds-web.wharton.upenn.edu/wrds/support/Data/_004Research%20Applications/

Output (who initiated)

Obs	SYMBOL	DATE	TIME	n_trades	price	mid_point2	buyer_init
1	DELL	19990104	9:45:03	2	74.688	74.656	1
2	DELL	19990104	9:45:04	3	74.685	74.656	1
3	DELL	19990104	9:45:05	2	74.656	74.656	-1
4	DELL	19990104	9:45:07	1	74.625	74.656	-1
5	DELL	19990104	9:45:08	4	74.656	74.656	1
6	DELL	19990104	9:45:10	4	74.680	74.656	1
7	DELL	19990104	9:45:11	6	74.629	74.656	-1
8	DELL	19990104	9:45:12	1	74.625	74.656	-1
9	DELL	19990104	9:45:13	1	74.688	74.656	1
10	DELL	19990104	9:45:14	3	74.638	74.656	-1
11	IBM	19990104	9:45:09	1	185.000	184.906	1
12	IBM	19990104	9:45:10	1	184.938	184.906	1
13	IBM	19990104	9:45:12	1	184.875	184.875	-1
14	IBM	19990104	9:45:23	1	185.000	184.875	1
15	IBM	19990104	9:45:29	1	185.000	184.875	1
16	IBM	19990104	9:45:56	1	185.000	184.938	1
17	IBM	19990104	9:45:58	1	185.188	184.938	1
18	IBM	19990104	9:46:14	1	185.125	184.938	1
19	IBM	19990104	9:46:30	1	185.125	185.094	1
20	IBM	19990104	9:47:21	1	185.125	185.063	1

wrds

Some issues

1) using primary market only

2) Use BBO or NBBO

3) Use 5-second before 1999 and 0-second after

4) Use NYSE/AMEX stocks only