

## IID Formula Note

Notation in this document:

- $P_k$  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) Algorithm:

Classify trade  $k$  as a buy if price  $P_k > M_t$ , as a sell if  $P_k < M_t$  and tick test is used when  $P_k = M_t$ , where  $t$  is the quote of the stock at the  $t$  second defined below.

- Note 1, Lee-Ready Same Second Quotes

In the Lee and Ready 1991 (LR) algorithm above, trade  $k$  and quote  $t$  take place in the same second.

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

- Note 2, Lee-Ready Previous Second Quotes

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  $P_k = M_{t-1}$

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

- Note 3, Lee-Ready Interpolated Second Quotes (Holden Jacobsen)

Interpolated Second is a timing method introduced by Holden and Jacobsen, "Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions", Journal of Finance, 2014.

Interpolated timing method solves the quote assigning problem of multiple trades populated in the same second quote. Please see Section VII of the paper for details.

- **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.

- **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

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

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

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

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

- [Formula 5](#), Simple Averaged Dollar Realized Spread

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

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

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

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

- [Formula 6](#), Simple Averaged Percentage Realized Spread

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

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

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

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

- [Formula 7](#), Simple Averaged Dollar Price Impact

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

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

$$PriceImpactDollar\_AVG_{i,T} = \frac{1}{N} * \sum_{k=1}^n Dollar\ Price\ Impact_k ,$$

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

- [Formula 8](#), Simple Averaged Percentage Price Impact

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

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

$$PriceImpactPct\_AVG_{i,T} = \frac{1}{N} * \sum_{k=1}^n Percent\ Price\ Impact_k ,$$

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

- [Formula 9](#), Value Weighted Average Price of Buy Trades

$$BuyVWAP_{i,T} = \sum_{k=1}^n w_k * P_k \mid k \in \{BUY_{i,T}\},$$

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

- [Formula 10](#), Value Weighted Average Price of Sell Trades

$$SellVWAP_{i,T} = \sum_{k=1}^n w_k * P_k \mid k \in \{SELL_{i,T}\},$$

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

- [Formula 11](#), Value Weighted Dollar Effective Spread

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

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

$$ESpreadDollar\_VW_{i,T} = \sum_{k=1}^n w_k \times Dollar\ Effective\ Spread_k ,$$

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 12](#), Share Weighted Dollar Effective Spread

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

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

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

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**

$$Percent\ Effective\ Spread_k = \frac{2D_k(P_k - M_k)}{M_k}$$

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

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

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

$$Percent\ Effective\ Spread_k = \frac{2D_k(P_k - M_k)}{M_k}$$

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

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

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

- [Formula 15](#), Share Weighted Dollar Realized Spread

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

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

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

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

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

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

$$RSpreadDollar\_VW_{i,T} = \sum_{k=1}^n w_k \times Dollar\ Realized\ Spread_k ,$$

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 17](#), Share Weighted Percentage Realized Spread

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

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

$$RSpreadPct\_SW_{i,T} = \sum_{k=1}^n w_k \times Percent\ Realized\ Spread_k ,$$

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

- [Formula 18](#), Value Weighted Percentage Realized Spread

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

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

$$RSpreadPct\_VW_{i,T} = \sum_{k=1}^n w_k \times Percent\ Realized\ Spread_k ,$$

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 19](#), Value Weighted Dollar Price Impact

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

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

$$PriceImpactDollar\_VW_{i,T} = \sum_{k=1}^n w_k \times Dollar Price Impact_k ,$$

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 20](#), Share Weighted Dollar Price Impact

$$Dollar Price Impact_k = 2D_k(M_{k+5} - M_k) ,$$

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

$$PriceImpactDollar\_SW_{i,T} = \sum_{k=1}^n w_k \times Dollar Price Impact_k ,$$

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

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

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

$$PriceImpactPct\_VW_{i,T} = \sum_{k=1}^n w_k \times Percent Price Impact_k ,$$

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

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

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

$$PriceImpactPct\_SW_{i,T} = \sum_{k=1}^n w_k \times Percent Price Impact_k ,$$

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](#), Second-by-second Intraday Volatility (trade based)

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  $Ret_{i,t} = Ln \frac{P_{k',t}}{P_{k,t-1}}$ .

- [Formula 24](#), Second-by-second Intraday Volatility (quote based)

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  $t$ .

- [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)

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

where  $VAR(Ret_{900t})$  is the variance of 15-minute log returns.

- [Formula 29](#), Variance Ratio 5 (30-min/2\*15-min)

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

where  $VAR(Ret_{1800t})$  is the variance of 30-minute log returns.

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

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

where  $SSqrtDVol = Sgn(\sum_{t-300}^t BuySHR - \sum_{t-300}^t SellSHR) \times \sqrt{|\sum_{t-300}^t BuySHR - \sum_{t-300}^t SellSHR|}$ ,  $M_{i,t} = \frac{B_{i,t} + A_{i,t}}{2}$  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:

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

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

- [Formula 32](#), Herfindahl Index

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



- **IID General Cleaning Filters:**

1. Eliminate testing symbols: 'TEST', 'TESTA', 'TESTB', 'TESTC', 'TESTD', 'TESTE', 'TESTG', 'TESTL', 'TESTM', 'TESTN', 'TESTP', 'TESTR', 'TESTV', 'ZVZZT', 'ZWZZT', 'ZXZZT', 'ZJZZT', 'ZYSTF', 'ZYSZZ', 'CJZZT', 'TJZZT', 'OJZZT', 'ATEST', 'AQZZT', 'EPZZT', 'OJZZT', 'IVZZT', 'ZBZZT', 'ZAZZT', 'ZYYZZ', 'ZYSBB', 'CMVTV', 'CNSIV', 'ZXYZ.', 'ZXYZ.A', 'ZXZ@', 'ZXZZ#', 'ZXZZ\*', 'ZXZZ+', 'ZXZZ^', 'ZXZ~', 'ZZO-A'.
2. Set the quotes with abnormal modes to missing: mode in (5, 16, 21, 22, 26 ).
3. One-Sided bid quotes (Bid>0 and Ask=0) are allowed to enter the NBB, but the ask will be set so that it won't enter the NBO; One-sided ask quotes (Bid=0 and Ask>0) are also allowed to enter the NBO, but no adjustment is necessary as the bid is already set to the extreme value 0.
4. Quotes with abnormally large spreads (greater than \$5) are set to missing, bid>0 and bid-ask>\$5.
5. Withdrawn quotes are set to missing values. This is when an exchange temporarily has no quote, as indicated by quotes with price or depth fields containing values less than or equal to 0 or equal to '!'. They are NOT deleted, because that would incorrectly allow the prior quote from that exchange to enter the NBBO. This filter must come last, see discussion in Holden and Jacobsen (2014).
6. The bid or ask at 4pm and dynamic trading time have to be non-missing. And at least there is one trade of the day.