The Accuracy of Trade Classification Rules for the Warsaw Stock Exchange

Paweł Miłobędzki[1], Sabina Nowak[2]

Abstract

We evaluate the accuracy of five trade classification rules for the Warsaw Stock Exchange: tick, reverse tick, quote, Lee and Ready and Ellis, Michaely and O'Hara. In doing so we use the transaction data on stocks from the large cap WIG20 index from the period May-September 2017. We find that the quote rule correctly classifies 100% of transactions initiated by buyers and sellers. Almost the same excellent job does the Lee and Ready rule. The Ellis, Michaely and O'Hara rule is less successful albeit its success rate exceeds 95% of the transactions assigned to both sides. The tick and the reverse tick rules exhibit a very low accuracy. The tick rule correctly classifies only 25.35% of transactions initiated by buyers and 25.95% of transactions initiated by sellers. The reverse tick rule performs even worse classifying as much as 16.66% and 16.67% of such transactions accordingly. The reason for their low accuracy is that the stock prices remain unchanged at the WSE at about 70% of all transactions. We also show that in case both classification rules are modified to account for either the preceding or the following transactions price changes their accuracy significantly increases.

Keywords: *accuracy of trade classification rules, market microstructure, Warsaw Stock Exchange*

JEL Classification: G10, G14, G15

DOI: 10.14659/SEMF.2018.01.32

1 Introduction

Information on parties to a trade who initiate transactions in financial markets plays an essential role in analysing the intraday price formation within the trade indicator models setup (Hagströmer et al., 2016). It is also helpful for determining the information content of trades, the order imbalance, the inventory accumulation of liquidity providers, the price impact of large transactions as well as the effective spread (Ellis et al., 2000). In case such information is not available which nowadays prevails because most public databases do not contain initiator flags, trade initiators and a trade direction are to be inferred using trade classification rules. These commonly employed in the empirical work include that of the tick (T), the reverse tick (RT), the quote (Q), the Lee and Ready (1991) (LR) as well as the Ellis et al. (2000) (EMO) (see Table 1 for their description). Their accuracy, albeit relatively high, varies across mature markets depending upon the trading price, the trade size and the time from previous trade, being rather low for short sales and trades inside the quotes (see Ellis et al., 2000; Odders-

[1] Corresponding author: University of Gdańsk, Department of Econometrics, ul. Armii Krajowej 101, 81-824 Sopot, Poland, e-mail: pawel.milobedzki@ug.edu. pl.

[2] University of Gdańsk, Department of Econometrics, ul. Armii Krajowej 101, 81-824 Sopot, Poland, e-mail: sabina.nowak@ug.edu.pl.

White, 2000; Finucane, 2000; Savickas and Wilson, 2003; Chakrabarty et al., 2007; Asquith et al., 2010; Rosenthal, 2012). To the best of our knowledge nothing has been known on the issue for emerging and developing markets so far except from the Taiwan Stock Exchange and the exchanges in Istambul and São Paulo (see Lu and Wei, 2009; Aktas and Kryzanowski, 2014; Perlin et al., 2014).[3]

Table 1. Trade classification rules.

Rule	Definition
TR	Current trade is a buy (sell) if its price is above (below) the closest different price of a preceding trade.
RT	Current trade is a buy (sell) if it is followed by a trade with a lower (higher) price.
QR	Current trade is a buy (sell) if its price is above (below) the mid-point of the bid and ask spread (mid-point price). Trades executed at the mid-point price are not classified.
LR	Current trade is a buy (sell) if its price is above (below) the mid-point price (as the QR). In case the trade price is equal to the mid-point price, the TR applies.
EMO	Current trade is a buy (sell) if it is executed at the bid (ask). In case it is executed at in between the quotes, the TR applies.

The aim of the paper is to overcome this deficiency for the WSE. To this end we use the transaction data on stocks from the large cap WIG20 index from the period 2 May-29 September 2017. The data come from Thomson Reuters.[4] We examine the stocks in the WIG20 because they account for approximately 80% in worth of all session trades in the analysed period. The period itself includes two quarterly revisions being held on 16 June and 15 September. Neither of them resulted in stock entries/exits into/from the index. The transaction data are stamped to the nearest millisecond. Each transaction record includes information on the trade price and the volume accompanied by the best bid and ask, the size of bid and ask

[3] Olbryś and Mursztyn (2015) estimated the fraction of trades classified as buys and sells at the WSE by the T, the Q, the LR and the EMO rules but they did not check for the accuracy of classification.

[4] The data are collected from the Thomson Reuters Eikon 4 database under the partnership agreement between the University of Gdańsk and the Thomson Reuters company. The system enables to download the data on trade no older than 3 months.

and the trade flag. The latter enables us to recognize the transaction type (regular, auction, cross, etc). Knowing that we identify a particular transaction as being initiated by a buyer (seller) if the resulting trade price is equal to the best ask (bid) or greater (lower) than that. Our definition of the trade initiator differs from those of Lee and Radhakrishna (1996) and Odders-White (2000) who recognize an investor as the initiator if he (she) places either a market order or places his (her) order chronologically last. We take into account only the normal trade, normal price and regular order transactions and exclude those of the open price, auction trade, trading at last, short sales as well as the cross ones. In the result our data set consists of 4658470 trades.

We find that the Q rule correctly classifies 100% of transactions initiated by the buyers and the sellers. Almost the same excellent job does the LR rule. The EMO rule is less successful albeit its success rate exceeds 95% of the transactions assigned to both sides. The T rule correctly classifies 25.35% of transactions initiated by the buyers and 25.95% of transactions initiated by the sellers. The RT rule performs even worse classifying 16.66% and 16.67% of such transactions accordingly. Their low accuracy is due to the fact that in the analysed period the stock prices at the WSE remain unchanged at about 70% of all transactions. In case we modify the T and the RT rules to account for either the preceding or the following transactions price changes their accuracy increases to reach 67.59% at most.

The remainder of the paper proceeds as follows. In Section 2 we sketch the present state of the art in examining the accuracy of common classification rules. In Section 3 we explain in more detail the nature of data we use in the empirical work and report on the accuracy of rules in question and their modification. The last Section briefly concludes.

2 Short Review of the Literature on the Classification Rules Accuracy

We gather the results of previous research on the accuracy of classification rules in Table 2.[5] Most of the findings invoked therein refer to the U.S. markets. Of a particular interest are those obtained for assets on data sets exhibiting a more detailed information on quotes and trades and referring to the multiplicity of rules.

Ellis et al. (2000) used the data on 313 newly traded NASDAQ stocks between 27 September 1996 and 29 September 1997 which contained the bid quote, the ask quote, the price, the trade volume, a trader identity code and a buy/sell indicator. By using the latter two they could determine whether a trade was a buy or a sell and documented 77.7%, 76.4%, 81.1%

[5] We report only on research performed on stocks and options written on them. The full list of papers regarding other assets is available to concerned readers on a request.

and 81.9% accuracy for the T, the Q, the LR and the EMO rules, respectively. They also find that all rules perform rather poorly in case of trades executed inside the quotes, large trades, trades during high volume periods and Electronic Communications Network trades.

Finucane (2000) based his research on the TORQ database which contained information on quotes, orders and trade direction for the sample of 144 stocks randomly selected from a size-stratified population of NYSE stocks for the three-month period November 1990 through January 1991. He compared the actual trade direction with that predicted by the T and the LR classification rules and found out that both methods were surprisingly about the same correctly identifying trade direction between 83% to 84% of the time. He suggested that the similar performance of both rules was due to the significant sample proportion of cross and stopped orders and trades between the quotes that might lower the accuracy of LR algorithm.

Lee and Radhakrishna (2000) on the same database noticed that approximately 40% of reported trades could not be unambiguously classified as either buys or sells mainly due to the presence of cross and stopped orders. Nevertheless 84.0% and 93.0% out of those classified agreed with the TORQ classification while identified with the use of the Q and the LR rules. The similar conclusions were reached by Odders-White (2000) who estimated the performance of the T, the Q and the LR rules at 78.6%, 74.9% and 85.0%, respectively. She also revealed that the rules systematically misclassified trades inside the quotes, small trades as well as trades in large or frequently traded stocks. Chakrabarty et. al (2007) concluded on trade classification rules accuracy using two samples consisted of 750 NASDAQ stocks traded on the INET ECN, the April and between April and June 2005 ones. The data contained information on the order, the trade history as well as on the buy/sell indicator. The major finding was that applying all rules to both samples resulted in almost equal success rate estimates around 74-77%. Rosenthal (2012) restricted attention to stocks included in the Russel 1000 large caps and 2000 small cap indices. His dataset covered transactions across 2836 stocks from three primary U.S. markets (AMEX, NASDAQ and NYSE) on two first trading days of December 2004. He showed that the introduction of delay models for estimating quotes and the modelling approach to trade classification led to a 1-2% improvement of the T, the LR and the EMO rule success rates in comparison to current methods.

Chakrabarty et al. (2015) using the true trade classification for equity transactions derived from the NASDAQ's Total View-ITCH order book and quotes and trades data from the Daily TAQ database provided by the NYSE showed that the success rate for the T and the LR rules were around 90-92%, but the latter rule was slightly more accurate across all strata (time, volume and trade bars for large, medium and small caps).

Table 2. Trade classification rules accuracy – summary of the literature review.

Study	Market, Instruments, Period	Rule			
		T	Q	LR	EMO
Ellis et al. (2000)	NASDAQ, 313 stocks, 96M9-97M9	77.7	76.4	81.1	81.9
Finucane (2000)	NYSE, 144 stocks, 90M11-91M1	83.0		84.4	
Lee and Radhakrishna (2000)	NYSE, 144 stocks, 90M11-91M1		84.0	93.3	
Odders-White (2000)	NYSE, 144 stocks, 90M11-91M1	78.6	74.9	85.0	
Savickas and Wilson (2003)	CBOE, 826 options, NASDAQ/OTC stocks, 95M7-95M12	59.4	82.8	80.1	76.5
Chakrabarty et. al (2007)	NASDAQ, 750 stocks, 05M4	75.6		75.8	76.8
	NASDAQ, 750 stocks, 05M5-M6	75.4		74.4	75.8
Asquith et al. (2010)	NASDAQ, 100 stocks, 05M3, M6, M12	41.3[s] 60.3[b]	37.5[s] 61.0[b]	39.1[s] 62.9[b]	
	NYSE, 100 stocks, 05M3, M6, M12	4.7[s] 98.9[b]	14.4[s] 83.6[b]	14.6[s] 88.0[b]	
Chakrabarty et al. (2012)	NASDAQ, 200 stocks, 05M6-M12			68.2[st] 69.2[lt]	
Rosenthal (2012)	NYSE, NASDAQ, AMEX, 2836 stocks, 04M12	66.2		71.7	72.8
Chakrabarty et al. (2015)	NASDAQ, 300 stocks, 11M5, M6, M7	90.8[i] 89.7[ii] 89.2[iii]		92.6[i] 91.6[ii] 91.2[iii]	
	NYSE, 300 stocks, 11M5, M6, M7	91.7[i] 89.0[ii] 87.8[iii]		93.4[i] 90.9[ii] 90.1[iii]	
Aitken and Frino (1996)	ASX, ECN, all stocks, 92M7-94M6	74.4			
Theissen (2001)	FSE, 15 stocks, 96M9-96M10	72.2	75.4	72.8	
Lu and Wei (2009)	TWSE, 684 stocks, 06M1-M6	74.2	92.8	96.5	95.0
Aktas and Kryzanowski (2014)	BIST, 30 stocks, 08M6-M12	90.4	95.0	96.4	86.9
Perlin et al. (2014)	BOVESPA, 15 stocks, 09M1-10M1	72.0			
Pöppe et al. (2016)	DB, 30 stocks, 12M10-M11	82.0		86.6	90.4

[b]Buyer initiated trades, [s]Seller initiated trades, [st]Short trades, [lt]Long trades, [i]One-hour time bar; [ii]100-trade bar; [iii]10000-volume bar. In case of Asquith et al. (2010) the maximum levels of accuracy are given.

Finally Savickas and Wilson (2003) compared the ability of the trade classification rules in question to classify options trades at the CBOE. Their dataset reported the trade direction on options underlain by 826 assets, mainly stocks from the NYSE/AMEX and the NASDAQ/ OTC. They estimated the rule success rate ranging from 59.4% (T rule) to 82.8% (Q rule). More interestingly, all rules happened to perform very poorly for the index options. The main source of their misclassification were outside-quote and reversed-quote trades.

The general conclusion stemming from the short review of accuracy rules performance on the U.S. markets is threefold. First, the estimates of accuracy apart from those reported in Asquith et al. (2000) for short sales range from mediocre to marvellous. Second, the LR rule and (or) the EMO rule perform better than their T and (or) Q counterparts on the NASDAQ, the NYSE and the AMEX, but on average the difference in their accuracy is rather slight (see Ellis et al., 2000; Finucane, 2000; Lee and Radhakrishna, 2000; Odders-White, 2000; Chakrabarty et al., 2007, 2015; Rosenthal, 2012). Third, trades in between the quotes and cross orders are deemed to have been the main sources of misclassification. The same applies to other international markets: the Deutsche Börse (Pöppe et al., 2016) and the Taiwan Stock Exchange (Lu and Wei, 2009). However at the CBOE, in Frankfurt and Istanbul the Q and the LR rules perform the best (see Theissen, 2001; Aktas and Kryzanowski, 2014; Savickas and Wilson, 2003).

3 Data and Empirical Results

Our primary dataset covers transactions on stocks from the WIG20 in the period 24 April-5 October 2017. It consists from trades labelled with 10 different trade flags indicating possible transaction types being executed at the WSE albeit it is the ‘Normal Trade, Normal Price, Regular Order’ trades that dominate the dataset.[6] Thus we clean it up dropping all residual flag trades as well as those of April and October to retain only those exhibiting the three-month regular trade in 20 largest cap stocks from May through September. Next we identify the trades as buys, sells and those executed inside the quotes (closer to the best ask/bid and not identified – traded at the mid-point price) comparing their trade prices with the best asks

[6] They constitute 96.63% of the dataset. The relevant frequency statistics are available to concerned readers on a request.

and the best bids that follow. We stack the results of cleaning up and identification of trades (trade initiators) in Table 3. They indicate that of all trades the buys at the best ask and the sells at the best bid prevail adding up to 99.35% of all cleaned trades. A tiny rest mainly consists of all inside the quotes trades including those not identified and those closer to the best ask and closer to the best bid. The fractions of buys above the best ask and sells below the best bid are residual.

Table 3. Trades included in the cleaned dataset by their type.

Trade type	No. of trades		Fraction	
Buy – Above Best Ask	1859		0.04	
Buy – At Best Ask	2323008	2324867	49.87	49.91
Inside – Closer to Best Ask	11756		0.25	
Inside – Not Ident	822	26366	0.02	0.56
Inside – Closer to Best Bid	13788		0.30	
Sell – At Best Bid	2305169		49.48	
Sell – Below Best Bid	2068	2307237	0.04	49.53
All trades		4658470		100.00

Table 4. Classification success rates for the T, the RT, the Q and the LR rules.

Trade type	T	RT	Q	LR	EMO
Buy – Above Best Ask	22.22	35.56	100.00	100.00	22.22
Buy – At Best Ask	25.35	16.64	100.00	100.00	95.82
Inside – Closer to Best Ask	20.58	15.24	100.00	100.00	20.58
Inside – Not Ident	87.10	66.79	100.00	87.10	87.10
Inside – Closer to Best Bid	21.78	12.98	100.00	100.00	21.78
Sell – At Best Bid	25.96	16.66	100.00	100.00	95.79
Sell – Below Best Bid	32.74	27.13	100.00	100.00	32.74
Buy	25.35	16.66	100.00	100.00	95.76
Inside	23.28	15.66	100.00	99.60	23.28
Sell	25.97	16.67	100.00	100.00	95.73
All trades	25.64	16.66	100.00	100.00	95.34

The results in Table 4 lay groundwork for the accuracy comparison of the rules in question. The accuracy statistics derived for all trades show that the Q rule performs the best correctly identifying all trades. The second is the LR rule which does almost the same excellent job. The next is the EMO which misclassifies only 4.66% of all trades. The remaining rules perform very poorly. The T rule misclassifies 74.36% of all trades while the RT rule misclassifies as much as 83.34% of them.

The accuracy statistics derived solely for buys, sells and inside the quotes trades indicate that the LR rule misclassifies only 0.4% of the latter. The EMO is slightly worse misclassifying 4.24% of buys, 4.27% of sells and 76.72% of inside the quotes trades. The accuracy of other rules for buys, sells and inside the quotes trades is about the same as their accuracy for all trades.

Table 5. Classification success rates for the modified T and RT rules.

Trade type	T1	T2	T3	T4	T5	RT1	RT2	RT3	RT4	RT5
Buy – Above BA[a]	31.79	36.47	39.27	40.77	41.64	51.86	59.33	64.17	67.24	69.34
Buy – At BA	40.62	50.55	57.45	62.45	66.18	26.26	32.40	36.63	39.70	42.00
Inside – Closer to BA	33.07	40.90	46.87	50.88	53.99	22.58	26.62	29.13	30.74	31.77
Inside – Not Ident	77.13	69.46	63.14	57.91	54.50	48.66	37.47	28.95	23.97	20.68
Inside – Closer to BB[b]	35.05	44.22	50.46	55.03	58.38	19.18	22.86	25.18	26.59	27.47
Sell – At BB	41.57	51.78	58.82	63.85	67.59	26.45	32.65	36.90	39.97	42.26
Sell – Below BB	49.85	59.04	64.94	69.29	72.10	39.51	46.23	50.39	53.29	55.27
Buy	40.61	50.54	57.43	62.43	66.16	26.28	32.42	36.65	39.72	42.03
Inside	35.48	43.53	49.26	53.27	56.30	21.62	24.99	27.06	28.36	29.18
Sell	41.58	51.79	58.82	63.86	67.59	26.46	32.66	36.92	39.98	42.27
All trades	41.06	51.12	58.07	63.09	66.81	26.35	32.50	36.73	39.78	42.07

[a]BA – best ask, [b]BB – best bid

The poor performance of the T and the RT rules (and the EMO for inside the quotes trades) is due to that in the period in question prices at the WSE remain unchanged at about 70% of all trades. But in case sequences of two and more consecutive trades are considered the fraction of them for which prices remain unchanged dramatically decreases. That is why we propose a modification of the T and the RT rules to account for either the preceding or the following transaction price changes. Their performance for different sequence lengths up to five is given in Table 5. In what follows the accuracy of both classification rules increases with the increasing length of sequence considered to reach at most 66.81% for the T5 rule. For all trades the estimates of accuracy statistics for the T rule are superior over those for the

RT rule. The difference in classification success rates ranges from 14.71% (R1 vs. RT1) to 24.74% (R5 vs. RT5). Almost the same applies to the buys, the sells and the trades executed inside the quotes. Nevertheless the accuracy rates for the modified RT rules are rather at unacceptable levels (below 50%). The accuracy of our classification results is comparable to those of Lu and Wei (2009), Aktas and Kryzanowski (2014), and Perlin et al. (2014) for TWSE, BIST and BOVESPA respectively, except for the T rule.

Conclusions

We estimate the accuracy rates of five common classification rules (T, RT, Q, LR, EMO) using the tick data on 20 large cap stocks listed at the WSE in the period May-September 2017. We find that the Q rule performs the best correctly identifying 100% of trades. The second is the LR rule which misclassifies only 0.4% of the inside the quotes trades. The next is the EMO misclassifying as much as 4.66% of all trades. The T and the RT rules perform very poorly misclassifying 74.36% and 83.34% of them, respectively. The reason is that in the analysed period prices at the WSE remain unchanged at about 70% of all trades. In case we modify the T and the RT rules to account for either the preceding or the following transactions price changes their accuracy increases to reach at most 66.81%. For all trades the estimates of accuracy statistics for the T rule are superior over those for the RT rule.

References

Aitken, M. & Frino, A. (1996). The accuracy of the tick test: Evidence form the Australian Stock Exchange. *Journal of Banking & Finance*, *20*, 1715-729.

Aktas, O. U. & Kryzanowski, L. (2014). Trade classification accuracy for the BIST. *Journal of International Financial Markets, Institutions & Money*, *33*, 259-282.

Asquith, P., Oman, R. & Safaya, C. (2010). Short sales and trade classification algorithms. *Journal of Financial Markets*, *13*(1), 157-173.

Chakrabarty, B., Li, B., Nguyen, V. & Van Ness, R. A. (2007). Trade classification algorithms for electronic communications network trades. *Journal of Banking and Finance*, *31*(12), 3806-3821.

Chakrabarty, B., Moulton, P. C. & Shkilko, A. (2012). Short sale, long sale, and the Lee-Ready trade classification algorithm revisited. *Journal of Financial Markets*, *15*(4), 467-491.

Chakrabarty, B., Pascual, R., & Shkilko, A. (2015). Evaluating trade classification algorithms: Bulk volume classification versus the tick rule and the Lee-Ready algorithm. *Journal of Financial Markets, 25*, 52-79.

Ellis, K., Michaely, R. & O'Hara, M. (2000). The Accuracy of Trade Classification Rules: Evidence from Nasdaq. *Journal of Financial and Quantitative Analysis, 35*(4), 529-551.

Finucane, T. J. (2000). A Direct Test of Methods for Inferring Trade Direction from Intra-Day Data. *Journal of Financial and Quantitative Analysis*, *35*(4), 553-576.

Hagströmer, B., Henricksson, R. & Nordén, L. L. (2016). Components of the Bid-Ask Spread and Variance: A Unified Approach. *Journal of Futures Markets*, *36*(6), 545-563.

Lee, C. M. C. & Radhakrishna, B. (2000). Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets*, *3,* 83-111.

Lee, C. M. C. & Ready, M. J. (1991). Inferring Trade Direction from Intraday Data. *Journal of Finance*, *46*(2), 733-746.

Lu, Y. C. & Wei, Y. C. (2009). Classification of trade direction for an equity market with price limit and order match: evidence from the Taiwan stock market. *Investment Management and Financial Innovations*, *6*(3), 135-147.

Odders-White, E. R. (2000). On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets*, *3,* 259-286.

Olbryś, J. & Mursztyn, M. (2015). Comparison of selected trade classification algorithms on the Warsaw Stock Exchange. *Advances in Computer Science Research, 12*, 37-52.

Perlin, M., Brooks, C. & Dufour, A. (2014). On the performance of the tick test. *Quarterly Review of Economics and Finance, 54*, 42-50.

Pöppe, T., Moos, S. & Schiereck, D. (2016). The sensitivity of VPIN to the choice of trade classification algorithm. *Journal of Banking & Finance*, *73*, 165-181.

Rosenthal, D. W. R. (2012). Modeling Trade Direction. *Journal of Financial Econometrics*, *10*(2), 390-415.

Savickas, R. & Wilson, A. J. (2003). On Inferring the Direction of Option Trades. *Journal of Financial and Quantitative Analysis*, *38*(4), 881-902.

Theissen, E. (2001). A test of the accuracy of the Lee/Ready trade classification algorithm. *Journal of International Financial Markets, Institutions and Money*, *11*(2), 147-165.