

Short sales and trade classification algorithms

Paul Asquith*, Rebecca Oman, Christopher Safaya

M.I.T. Sloan School of Management and NBER, USA

Available online 30 September 2009

Abstract

This paper demonstrates that short sales are often misclassified as buyer-initiated by the Lee–Ready and other commonly used trade classification algorithms. This result is due in part to regulations which require that short sales be executed on an uptick or zero-uptick. In addition, while the literature considers “immediacy premiums” in determining trade direction, it ignores the often larger borrowing premiums that short sellers must pay. Since short sales constitute approximately 30% of all trade volume on U.S. exchanges, these results are important to the empirical market microstructure literature, as well as to measures that rely upon trade classification, such as the probability of informed trading (PIN) metric.

© 2009 Elsevier B.V. All rights reserved.

JEL classifications: G10; G18

Keywords: Market microstructure; Trade classification; Short sales

The Lee–Ready algorithm, defined in [Lee and Ready \(1991\)](#), is widely used to establish the direction of an equity trade. That is, it classifies whether the transaction is initiated by a buyer or seller. The issue of trade direction is important in the market microstructure literature, particularly for information-based trading models. Commercially available trade and quote data, however, do not include the direction of trades. As a result, empiricists use the Lee–Ready and other commonly accepted algorithms, such as the tick test and quote test, to assign trade direction. As of October 12, 2009 the Web of Knowledge lists 350 citations for [Lee and Ready \(1991\)](#), and Google Scholar lists 1008.

*Corresponding author. Tel.: +1 617 253 7177; fax: +1 617 253 0603.

E-mail addresses: pasquith@mit.edu (P. Asquith), oman@alum.mit.edu (R. Oman), csafaya@alum.mit.edu (C. Safaya).

This paper applies the three algorithms to short sale trades and finds that all three classify the majority of short sales as buyer-initiated.¹

The tick test classifies a trade as a buy (sell) if the trade price is higher (lower) than the previous trade. If the current and previous trade prices are the same, the trade is classified by the next previous trade. Because the test depends on data from previous trades, transactions at the beginning of the day are not classified. The quote test classifies a trade as a buy (sell) if the trade price is closer to the ask (bid). This test, however, is unable to classify those trades priced at the midpoint between the bid and ask. The Lee–Ready algorithm combines the trade and quote tests to establish trade direction. It first classifies a trade based on the quote test, then applies the tick test to the trades priced at the midpoint between the bid and ask.

Although there is no theoretical framework justifying the Lee–Ready and other algorithms, empirical studies have typically concluded that these methods are reasonably accurate. The Lee–Ready algorithm is usually reported as the most accurate, though the degree of accuracy varies, ranging from 72% to 93%, depending on the study. Tests of these algorithms, however, run into the same problem that makes them necessary. That is, there is no good source of trade data that identifies transactions as buyer- or seller-initiated. Those papers that do test the accuracy of these methods typically use datasets that are limited in time, have a small sample size, or are from non-U.S. exchanges. Further, since these limited datasets do not unambiguously identify trades as buyer- or seller-initiated, researchers are forced to rely upon rules of inference. However, all the rules of inference that are employed only apply to single-sided market or executable limit orders, making large portions of the datasets unusable. In addition, all of the current tests of Lee–Ready algorithm occurred before decimalization in 2001.² Decimalization has led to narrower bid–ask spreads, which may make trade classification more difficult.

The trade direction literature identifies the initiating party by focusing on two characteristics. First, the initiator is defined as the last party to enter a transaction. Lee and Radhakrishna (2000), Odders-White (2000), and Ellis et al. (2000) use this definition. The rationale is that the first party does not select a price that results in immediate execution. Rather, the first party acts as a liquidity provider at their chosen price. Second, the initiator is defined as paying an “immediacy premium” for rapid execution of the trade, a point that Lee and Radhakrishna (2000) and Odders-White (2000) make. In this paper, we consider both definitions as they apply to short sales.

This paper applies the three commonly used trade classification algorithms to a new dataset of short sale transactions for stocks on the NYSE and NASDAQ in 2005. Short sales are unique because a short seller must locate and borrow, at a positive interest cost, the stock being sold. Short sales are also highly regulated and are subject to the uptick rule on the NYSE and the bid–price rule on the NASDAQ. For NYSE stocks, the uptick rule requires that a short sale must be executed on an uptick or zero-uptick. For NASDAQ stocks, the bid–price rule requires that a short sale must be executed above the previous inside bid.

¹We also used the Ellis, Michaely, and O’Hara trade classification algorithm. The EMO results are similar in size, direction, and statistical significance to all the results reported in this paper. However, since EMO is less commonly used than the others, to save space, we do not report those results in this paper.

²Boehmer et al. (2007) use post-decimalization data, but their focus is the impact trade misclassification has on the probability of informed trading (PIN) estimation.

In this paper, we find that the trade classification algorithms overwhelmingly classify short sales, a unique subset constituting almost 30% of trading volume, as buyer-initiated. We then address the question of whether short sales are properly classified. Analyzing the classification algorithms, we find that all three classify about two-thirds of the short sale trades in our sample as buyer-initiated. This is in part due to the uptick or bid–price rules mentioned above. In 2005, the SEC exempted a number of stocks from these rules in a pilot study. Testing these exempted securities, we find that short sales on pilot securities are classified by Lee–Ready as buyer-initiated 52% to 59% of the time depending on the exchange and time period.

We demonstrate that the uptick and bid–price rules cause many short sales to be improperly classified as buyer-initiated. These rules cause short sales to be executed at a price that is higher in the bid–ask spread. The Lee–Ready and other algorithms reflect this bias, regardless of the order in which each party entered the transaction. We then consider the size of a short seller’s execution costs relative to the immediacy premium. At a minimum, our findings are surprising, and they introduce some concern regarding the accuracy of existing trade classification algorithms. In addition, we raise concerns about applying the trade classification algorithms to all trades in the TAQ database.

The remainder of the paper is organized as follows. Section 1 reviews previous empirical tests of the Lee–Ready and other trade direction algorithms. Section 2 discusses our dataset and methodology. Section 3 presents our results. Section 4 discusses implications of applying the Lee–Ready algorithm to trades that contain short sales. Section 5 concludes.

1. Literature review

In their 1991 paper defining the algorithm, Lee and Ready examine a sample of trades for 150 NYSE firms in 1988. They first criticize the tick test, stating that its primary limitation is its reliance on previous trades, which may not be current. Since trades may occur several minutes or more apart, they argue that quotes are a better indicator of current market conditions. However, the paper also identifies two potential problems with the quote test. The first is that the prevailing quote at the time of a trade may not be that trade’s associated quote. Specifically, they find that quote changes resulting from a trade may be recorded ahead of that trade. The Lee–Ready algorithm addresses this problem by adjusting the quote data to exclude those quotes that occur less than 5 seconds before the current trade. The second problem with the quote test is that trades at the midpoint between the bid and ask are not classifiable. For these trades, the Lee–Ready algorithm relies on the tick test.

Lee and Radhakrishna (2000) test the Lee–Ready algorithm using the TORQ dataset. This database is comprised of 144 NYSE stocks trading during the three-month period from November 1990 through January 1991. Trade direction is inferred by tracing transactions to the originating party through the use of the NYSE system order file. In an attempt to ensure that each observation has only one active (initiating) party, the authors eliminate stopped market orders, market “crosses,” and the pairing of market orders with executable limit orders. Removing these order types eliminates approximately 40% of the trades in the TORQ dataset. Lee and Radhakrishna (2000), using a sub-sample of 15 stocks and 129,700 trades, find that the Lee–Ready algorithm has a 93% overall success rate for the remaining single participant trades. The success rate is highest for those trades

at the bid or ask, with 98% of trades correctly classified. The method is less accurate, 76%, for midpoint trades classified by the tick test.

Finucane (2000) uses the TORQ dataset to test the Lee–Ready algorithm. He compares it to the tick test and revisits the question of time-matching trade and quote data. His sample consists of 337,667 trades for all 144 firms, with 25% of the original transactions eliminated. Much like Lee and Radhakrishna (2000), Finucane only includes those trades that contain a market order on at least one side of the trade. He tests the algorithms using quotes adjusted according to Lee and Ready's (1991) original 5-Second rule, as well as unadjusted quotes. The results are similar for both, indicating that delayed trade timestamps may not pose as large a problem as Lee and Ready originally stated. Additionally, Finucane finds that the Lee–Ready algorithm has a success rate of 84.4%, while the tick test correctly classifies 83.0% of the trades in the sample.

Odders-White (2000) also uses the TORQ dataset to evaluate all three trade direction algorithms on a sample of 318,364 transactions. She reports that 25.1% of the observations in the database cannot be unambiguously classified as buyer- or seller-initiated. Odders-White infers the buy/sell intent of the initiating party on the basis of time order. She defines the initiating party as the last party, chronologically, to place an order. This assumption, she argues, is consistent with the idea that the initiating party is willing to pay a premium for immediate execution. She also notes that it is not possible to identify an initiating party when all parties to the trade are active (a crossed trade, for example). The paper finds that the tick test misclassifies 21.4% of the trades, the quote test misclassifies 9.1% of the trades, and the Lee–Ready algorithm incorrectly classifies 15.0% of the trades. Odders-White finds that trades within the spread, small trades and trades in large or frequently traded stocks are more likely to be misclassified. Odders-White also makes the point that while the TORQ database identifies order type, most other databases do not.

Ellis et al. (2000) use a proprietary NASDAQ dataset to analyze the tick test, quote test, and Lee–Ready algorithm. Their sample is comprised of 313 NASDAQ stocks trading from September 27, 1996 through September 29, 1997, and contains 2,433,019 trades. The authors establish trade direction using the “buy/sell indicator” included in their dataset, which identifies the reporting party. This indicator, in conjunction with trader identity codes included in the dataset, allows the authors to identify whether the initiating party is a market-maker, a broker, or a customer. They include only those trades that take place between market-makers and brokers/customers or trades between brokers and their customers. In the end, 24.6% of the sample is discarded as a result of ambiguity. Ellis et al. find that the tick test, quote test, and Lee–Ready algorithm correctly classify 77.7%, 76.4%, and 81.1% of the sample, respectively. As is the case with the other empirical tests, their paper finds significantly higher accuracy rates for trades at the bid and ask prices.

Theissen (2001) tests the accuracy of the Lee–Ready algorithm and the tick test using data from the Frankfurt stock exchange. The sample in his paper tracks 15 stocks from September 26, 1996 through October 25, 1996. Trade direction is inferred from the position taken by the stock's specialist. Theissen does not concern himself with the problems associated with order type since all trades in the sample include a market-maker. A sample of 9,124 transactions is used for the tick test, 72.2% of which agree with Theissen's classification of the trades. Theissen then tests the Lee–Ready algorithm using 9,449 observations and finds that it agrees with the specialist classification for 72.8% of the trades.

Aitken and Frino (1996) examine transactions on the Australian Stock Exchange (ASX) for the two-year period ending June 30, 1994. The paper evaluates the performance of only the tick test on its sample of 4,022,339 transactions. The authors infer trade direction on the basis of time priority. They consider the last party to enter or amend an order to be the initiating party. The paper finds that the tick test classifies 74.4% of the trades in the sample in a manner consistent with their buyer and seller assignments, but it finds that seller-initiated trades are more frequently misclassified than buyer-initiated trades. Aitken and Frino suggest that short sales may be to blame for this misclassification, noting that the ASX is subject to a tick rule similar to those used for many U.S. exchanges.³

2. Data and research design

In order to test the trade direction algorithms, we randomly select from CRSP 200 common stocks that were listed in 2005. One hundred of these stocks are taken from the NYSE and 100 from NASDAQ's National Market System. For our analysis, we examine the time periods March 1, 2005–March 31, 2005, June 1, 2005–June 30, 2005, and December 1, 2005–December 31, 2005.

After selecting our sample of stocks and choosing our time period, we then combine and analyze trade and quote information from two databases. The first is the TAQ database, which is commonly used in the empirical market microstructure literature. The dataset contains two files, one of which describes all trades occurring on each exchange, while the other describes the quotes. We merge the two files to obtain trade and quote data for each transaction in our sample. For each stock, the TAQ data provide a ticker symbol, exchange code, timestamp, bid and ask price and size (for quotes), and transaction price and size (for trades). Trade direction and order type, however, are not provided. For our sample stocks and time periods, the TAQ database contains 12,099,132 trades and 70,877,601 quotes. This represents 7,170,782,500 shares traded.

To establish which trades are short sales, we rely on a second, non-commercial database compiled by Paul Asquith, Andrea Au, and Parag Pathak, which contains all short sale transactions in 2005. SEC Regulation SHO went into effect at the beginning of 2005, requiring each exchange to post data describing individual short transactions. The data made available include the price, time, and size of all short sales, in addition to ticker symbols and exchange codes identifying the security and where it is traded. Their database aggregates this information across the nine major U.S. exchanges.

We then use the short sale database to identify which trades in our sample of 200 stocks are short sales and classify them as seller-initiated. There are 3,587,468 short trades in our sample during the relevant time periods. These short trades represent 29.7% of all trades for our stocks during the three months and 27.9% of all trading volume.⁴ Observations are matched by symbol, exchange, price, size, date, and time. For electronic transactions, times are recorded exactly and the matching is straightforward. Floor negotiated transactions, however, may receive a timestamp at the beginning, middle or end of the negotiation. As a

³Diether et al. (2009) also hypothesize that the uptick rule should increase the percentage of NYSE short trades above the bid–ask midpoint when using the Lee–Ready algorithm. Because of regulatory differences, they further hypothesize that the bid–price test will not have such an effect for NASDAQ short trades.

⁴This percentage is in line with that reported in the short sale literature. Asquith et al. (2006) show that for all common stocks during all of 2005, short sales are 28.1% of total trading volume. Diether et al. (2008) find a similar number of 26.7% for their sample of NASDAQ stocks.

Table 1
 Characteristics of the 200 randomly selected stocks in our sample.

	Number of trades	Number of shares traded	Price/ share	Number of quotes	Bid–ask range	Number of short trades	Number of shares shorted	% Shares sorted
March								
All	21,320	12,526,613	\$28.91	113,324	\$0.13	5,773	3,109,806	25.1
NYSE	26,519	15,782,125	\$34.38	105,057	\$0.04	6,830	3,369,002	23.3
NASDAQ	16,121	9,271,101	\$23.44	121,592	\$0.22	4,715	2,850,610	27.0
June								
All	19,402	11,973,618	\$28.44	116,246	\$0.12	5,844	3,355,520	26.6
NYSE	24,592	13,001,244	\$33.71	104,054	\$0.04	6,976	2,982,895	24.5
NASDAQ	14,212	10,945,993	\$23.16	128,438	\$0.20	4,711	3,728,145	28.7
December								
All	19,774	11,353,681	\$29.99	124,818	\$0.12	6,321	3,522,072	27.6
NYSE	26,091	12,798,697	\$35.17	105,425	\$0.04	8,051	3,435,775	26.6
NASDAQ	13,457	9,908,665	\$24.81	144,212	\$0.21	4,590	3,608,368	28.5

There are 100 NYSE stocks and 100 NASDAQ stocks for three months: March, June and December of 2005. The price per share is the average trading price during that month. The Bid–ask range is the average quote range per stock for the month. All numbers are per firm averages calculated by equally weighting each stock in the selected sub-samples.

result, a negotiated trade may have a timestamp in the SHO data that is not the same as the timestamp in the TAQ data. To correct for this, we allow a 1 second window around a given trade for the purpose of matching. For our sample stocks, 98.8% of the observations in our short sale database were exactly matched to transactions in the TAQ database, and 99.5% were matched within a 1 second window. The 0.5% not matched was eliminated from the sample.

Our dataset differs from those in the previous literature in a few key ways. The number of trades we examine is significantly larger than most of the prior tests of the Lee–Ready algorithm. In addition, our sample is from 2005 and only contains trades that took place after decimalization.⁵ Most importantly, all trades in our dataset are unambiguously classified as short or long trades.

An important consideration for classifying short sales is the effect of the uptick rule and bid–price rules. For example, because a non-pilot NYSE short sale must occur on an uptick or zero-uptick, it is not always possible to execute a short sale as a market order even if it is entered as one. Diether et al. (2009) posit that the uptick rule has a significant effect on how short sales are executed since the NYSE trade software adjusts short trades to ensure compliance with the uptick rule. This software effectively converts non-compliant market order short sales into limit orders. Pilot securities, however, are exempt from the uptick rule and face no such limitations. A higher percentage of pilot securities

⁵Bessembinder (2003a, b) shows that decimalization results in narrower bid–ask spreads, and Ellis et al. (2000) as well as Lee and Radhakrishna (2000) have shown that the trade direction algorithms are more accurate at the bid and ask than at the quote's midpoint. If narrower spreads result in prices being closer to the midpoint, classification could become more problematic. If narrower spreads result in less midpoint trades, then the classification could actually become more accurate. These hypotheses have never been tested empirically.

should, therefore, be classified correctly since pilot short sales are more likely to be executed as market orders.

Table 1 provides a descriptive overview of our sample and lists the per-stock average for number of trades, number of shares traded, price-per-share, number of quotes, bid–ask range, number of short trades, number of shares shorted, and percentage of short sales by month, as well as by exchange for the entire sample. The average number of shares traded is relatively constant across exchanges and time periods. The same holds for the percentage of short trades and the number of shares shorted. Almost 30% of all trades are short sales, which represents over 27% of shares traded by volume. The bid–ask range is much larger for NASDAQ securities than for NYSE securities, approximately \$0.21 and \$0.04, respectively. The number of quotes is approximately four times as large as the number of trades for NYSE stocks, and over nine times as large as the number of trades for NASDAQ stocks.

We make an additional adjustment to the combined dataset before testing the trade direction algorithms. As mentioned above, Lee and Ready (1991) only consider quotes that are at least 5 seconds older than a trade when testing their algorithm. Vergote (2005) revisits this problem for data on electronic exchanges. He finds that the introduction of electronic communications networks improves the accuracy of quote and trade time-stamps, and recommends using a 2 second delay. Finally, when assessing whether trades are buyer- or seller-initiated, Bessembinder (2003a, b) finds that no time adjustment is necessary when aligning the quotes and trades for NASDAQ and NYSE stocks. While matching quotes with trades, we adjust our dataset to test the 5 second delay, 2 second delay, and no delay.

After these adjustments, we classify the short sales in our sample as either buyer- or seller-initiated using the tick test, the quote test, and the Lee–Ready algorithm. For the tick test, we only consider previous trades from the same exchange as the current trade. This is consistent with NYSE rules that forbid using off-exchange trades to satisfy the uptick rule.⁶ We do the same for quotes, only using prior quotes from the same exchange, although there is no corresponding regulatory requirement. Finally, we include only those trades and quotes that take place during normal trading hours.

In addition to requiring exchanges to post data describing short sale activity, Regulation SHO began a pilot study in May 2005, which exempted short sellers in certain stocks from adhering to the uptick and bid–price rules. This regulatory change allows us to test the Aitken and Frino (1996) and Diether et al. (2009) conjectures that these particular trading restrictions have an effect on the classification algorithms. Our sample of 100 NYSE stocks contains 35 pilot securities and our sample of 100 NASDAQ stocks contains 22. This paper tests pilot and non-pilot securities separately.

Since the pilot study began in May 2005, the periods we have selected (March, June, and December, 2005) allow us to examine the effects of this regulatory change in two ways. First, we compare pilot and non-pilot firms during the pilot period. Second, we examine the pilot firms both before and after the pilot program was implemented. Finally, for our sample stocks and time periods, we examine all long trades in the TAQ dataset in order to establish a benchmark for comparison. Since we do not know whether long trades are

⁶Earlier studies do not discuss whether they include trades off the exchange. Trading securities off of their primary exchange is currently a common occurrence. In fact, for NYSE listed stocks in 2005, 23.4% of short trading volume takes place on other exchanges. Source: Asquith et al. (2006).

Table 2
Trade classification results using the tick test, quote test, and Lee–Ready algorithm.

Tick test				Quote test			Lee–Ready		
Seller-initiated (%)	Buyer-initiated (%)	Non-classifiable (%)		Seller-initiated (%)	Buyer-initiated (%)	Midpoint and non-classifiable (%)	Seller-initiated (%)	Buyer-initiated (%)	Non-classifiable (%)
Short Sale ($N = 3,587,468$) ^a									
5 s delay				27.0	68.5	4.5	28.2	71.5	0.3
2 s delay	24.3	75.5	0.3	26.9	69.5	3.6	27.9	71.8	0.2
No delay				30.3	58.2	11.4	33.4	66.5	0.1
Long trades($N = 8,511,664$) ^a									
5 s delay	57.8	42.1	0.1	53.9	40.9	5.3	56.8	43.0	0.2
2 s delay				54.3	41.4	4.4	56.7	43.1	0.1
No delay				46.8	40.0	13.2	54.8	45.1	0.1

The tick test, quote test, and Lee–Ready algorithm are used to classify trades as seller- or buyer-initiated for the sample of 100 NYSE and 100 NASDAQ stocks during March, June, and December 2005. The percentages are the number of trades classified as sells, buys, or not classifiable divided by the total number of trades. N is the total number of either short or long trades. The delays of 5second, 2second, and no delay define which quote is considered active at the time of the trade for the quote test and Lee–Ready algorithm.

^aThe seller- and buyer-initiated classification percentages for all classification algorithms and all possible delays for short sales in the top half of Table 2 are significantly different at the 0.1% level from the seller- and buyer-initiated classification percentages for long trades in the bottom half of Table 2.

buyer- or seller-initiated, we can only provide the percentage classified in each category by the trade algorithms.

3. Results

Table 2 presents the results from applying each algorithm to our entire sample of trades, and it shows that each algorithm overwhelming classifies short sales as buyer-initiated. The sample trades are first divided into short and long transactions. The quote test and Lee–Ready algorithm are evaluated using Lee and Ready’s 5 second quote matching delay, the 2 second quote matching delay suggested by Vergote, as well as no delay. Quote delays are irrelevant for the tick test since quotes are not used. The first column of the top panel in Table 2 shows that the tick test classifies only 24.3% of the short sale transactions in our sample as seller-initiated. Moreover, the quote test, using the 5second delay identifies 27.0% of the short trades as seller-initiated. The Lee–Ready algorithm, with a 5 second quote delay, classifies 28.2% of the short sales in our sample as seller-initiated.

The results for the 5second quote delay, 2second quote delay, and no delay are not materially different, although the no delay case identifies the highest percentage of short sales as seller-initiated. The results for long trades follow in the bottom panel. This panel shows that more long trades are classified as seller-initiated than short sales. Between 46% and 57% of the long trades are classified as seller-initiated, and between 40% and 46% are classified as buyer-initiated. Under each classification algorithm, and with all three possible

delays, the difference between the classification results of short sales and long trades is statistically significant at the 0.1% level or higher. That is, short sales have a significantly lower percentage of trades classified as seller-initiated than long trades.

A possible explanation for the overwhelming classification of short sales as buyer-initiated is that most of them are subject to either the uptick or bid–price rules. For example, since the uptick rule allows short trading on only an uptick or zero-uptick, short sales subject to the rule would be automatically classified incorrectly as buyer-initiated using the tick test. The quote test should also be affected by the uptick or bid–price rules. Floor traders treat short sales subject to the uptick rule as limit orders executable only at or above the last trade price. This behavior causes short sales to be executed at a higher price within the spread than they otherwise would be. Thus these regulations introduce a systematic bias, since any trades above the midpoint of the spread are classified by the quote test as buyer-initiated. Because the Lee–Ready algorithm is a combination of the tick and quote tests, it is prone to the same bias as those tests. The effect of the bid–price rule should be similar to that of the uptick rule, though not as severe since application of the bid–price rule is more flexible.

In Table 3, we test whether the misclassification of short sales is a result of the NYSE uptick rule and the NASDAQ bid–price rule by applying all three algorithms to pilot and non-pilot short trades separately. We also divide our sample into monthly time periods as a second test of whether the uptick rule affects the trade classification algorithms. Since our March sample is before the SEC’s pilot study, which began in May, we can test if pilot firms are classified differently before and after the uptick and bid–price rules were suspended. Finally, we divide the sample into NYSE and NASDAQ stocks, to see if the application of the uptick rule is different from the bid–price rule. Table 3 presents our results when using the 5 second quote delay. The results for the 2 second and no delay adjustments do not differ substantially but, just as in Table 2, the no delay case classifies the highest percentage of short sales as seller-initiated.

The NYSE results shown in the top half of Table 3 are dramatic. For the non-pilot NYSE sub-sample, 98.9%, 98.5%, and 95.0% of all short sales are assigned as buyer-initiated by the tick test in March, June, and December, respectively. The quote test and Lee–Ready algorithm classify over 80% of the non-pilot NYSE short sales as buyer-initiated in each month. The difference for pilot NYSE stocks is striking. For example, the Lee–Ready algorithm classifies only 11.8% of the non-pilot short sales as seller-initiated in June but classifies 40.6% of the pilot short sales as seller-initiated in the same month. The results are similar for December. The differences between pilots and non-pilots in June and December are statistically significant at the 0.1% level and indicate that the uptick rule causes a substantial amount of short sales to be misclassified as buyer-initiated when they are, in fact, seller-initiated.

The results for the three algorithms on NASDAQ short sales are not as one-sided, with approximately 60% of the non-pilot stock trades classified as buyer-initiated by the tick test, quote test and Lee–Ready algorithm for all three months. Even though the difference is not as large as for the NYSE trades, the percentage of NASDAQ pilot short sales classified as seller-initiated is higher than that of non-pilots in June and December. For example, the Lee–Ready algorithm classifies 41.6% and 46.8% of the pilot short sales as seller-initiated in June and December, but only 36.6% and 39.1% of the non-pilot short sales in the same months. The differences in classification percentages for June and December between pilot and non-pilot are statistically significant at the 0.1% level for all three algorithms.

Table 3
Short sale classification results using the tick test, quote test, and Lee–Ready algorithm.

	Number of trades (%)	Tick test			Quote test			Lee-Ready			
		Seller-initiated (%)	Buyer-initiated (%)	Non-classifiable (%)	Seller-initiated (%)	Buyer-initiated (%)	Midpoint and non-classifiable (%)	Seller-initiated (%)	Buyer-initiated (%)	Non-classifiable (%)	
NYSE											
March											
Non-pilot	445,812	0.9	98.9	0.2	12.1	83.3	4.5	12.2	87.6	0.2	
Pilot	237,223	0.4	99.3	0.2	12.6	82.8	4.5	12.7	87.1	0.2	
June											
Non-pilot ^a	427,211	1.3	98.5	0.2	11.7	83.6	4.6	11.8	88.0	0.2	
Pilot ^a	270,416	41.5	58.2	0.3	38.1	56.2	5.7	40.6	59.2	0.3	
December											
Non-pilot ^a	480,776	4.7	95.0	0.3	14.4	81.0	4.6	14.6	85.1	0.2	
Pilot ^a	324,353	44.4	55.3	0.2	41.6	52.8	5.7	44.2	55.6	0.2	
NASDAQ											
March											
Non-pilot	282,893	39.4	60.3	0.3	35.1	59.8	5.1	37.3	62.3	0.4	
Pilot	188,641	38.3	61.5	0.2	34.1	61.9	4.0	35.9	63.8	0.4	
June											
Non-pilot ^a	261,277	39.7	59.8	0.4	34.8	61.0	4.2	36.6	62.9	0.5	
Pilot ^a	209,820	42.7	57.0	0.3	40.4	56.5	3.1	41.6	57.9	0.5	
December											
Non-pilot ^a	244,878	41.3	58.2	0.5	37.5	58.9	3.6	39.1	60.5	0.4	
Pilot ^a	214,168	47.8	52.1	0.2	45.7	51.6	2.7	46.8	52.9	0.3	

The percentage of trades classified as seller-initiated, buyer-initiated, or non-classifiable by the tick test, quote test and Lee–Ready algorithm on the sample of short sales for 100 randomly-selected NYSE and 100 randomly-selected NASDAQ stocks during the months of June and December, 2005. The quote test and Lee–Ready algorithm use a 5 second quote delay. The sample is divided by whether the stock is included in the SEC Pilot study.

^aThe seller-initiated classification percentages for non-pilot short sales are significantly lower than the classification percentages for Pilot short sales in June and December for all three classification algorithms at the 0.1% level. This is true for both NYSE short sales in the top half in Table 3 and for NASDAQ short sales in the bottom half of Table 3. This is not true for March NYSE or NASDAQ short sales.

The pilot and non-pilot differences for June and December are further supported by examining the March results (before the uptick and bid–price rules were suspended) where there is no significant difference between pilot and non-pilot short sales. In March, the Lee–Ready algorithm classifies 12.7% of pilot NYSE short sales as seller-initiated and 12.2% of non-pilot NYSE short sales as seller-initiated. This difference is not significant.⁷ In addition, for NASDAQ short sales in March, the Lee–Ready algorithm actually classifies a higher percentage of non-pilot stocks (37.3%) as seller-initiated than of pilot stocks (35.9%). Thus, not only are the classification differences between pilot and non-pilot NYSE and NASDAQ stocks in June and December large and statistically significant, the differences in March are neither.

The results in Table 3 show that while both rules matter, the uptick rule has a larger percentage impact on the classification algorithms than the bid–price rule. It is also important to note that the seller-initiated classification rates for pilot securities in June and December, while higher than non-pilots, still classify less than half of the short trades in our sample as seller-initiated. Therefore, even without these trading restrictions, all three algorithms predominantly classify shorts sales as buyer-initiated.

Examining the natural experiment created by regulation SHO seems to demonstrate that the uptick and bid–price rules cause many short sales to be misclassified as buyer-initiated. Another possibility, however, is that the differences in pilot and non-pilot results can be explained by fundamental changes in trader behavior. That is, traders changed their trading strategies due to the elimination of the uptick and bid–price rules. We find this explanation for the significantly larger amount of seller-initiated short sales to be unlikely. This “Lucas Critique” of our results does not hold unless short sellers change the parameters of their trading model. We argue that the basic parameter for short sellers remains profit maximization before and after regulation SHO. Short sellers short stocks when they think the stock price is going down, just as traders buy when they think the stock price will go up. If regulation imposes a cost to that process, the economic implication is that there must be a greater gain to the short sale in order for a trader to undertake it. This in turn reduces the amount of short selling, but not the economics of which stocks are overpriced. Removing the regulation, which the pilot study does, does not change how short sellers identify which stocks are overpriced or their goal of profit maximization; it merely removes one cost.⁸

There are a number of recent papers that investigate this issue of whether short selling behavior changed as a result of the pilot study. The SEC’s *Economic Analysis* (2007) of the pilot program reported an increase in short sales volume of 2%, and concluded that removal of short sale restrictions did not result in any material adverse impact on either market quality or liquidity as measured by the relative effective quoted spread. Alexander and Peterson (2008) also test this conclusion using a sample of 418 pilot and non-pilot securities. They conclude, “the removal of price tests has had no deleterious effect on

⁷All statistical tests are robust to non-*iid* errors and all are corrected for stock and date dependence.

⁸Examining a regulatory regime change is a common way to test for the effects of regulation. In addition, a broad view of the “Lucas Critique” would also apply to the change to decimalization in 2001. Since decimalization leads to lower bid–ask spreads and since trade classification algorithms depend on where in the bid–ask spread a trade takes place, it could be argued (although we do not) that this regulatory change invalidates trade classification research using data after 2001.

trader behavior and has not led to a decrease in market quality.” Alexander and Peterson do find, however, that the elimination of the uptick rule results in improved order execution speed and marginally lower execution prices. Diether et al. (2008, 2009) find that, for a sample of pilot and non-pilot securities, the total increase in short sale volume is approximately 2%, the same change reported by the SEC’s economic review. All of these results are consistent with the pilot study removing a regulatory cost to short selling, but not changing the underlying decision model for short sellers.⁹

Examining Table 1, we find that short sales increase by approximately 1.5% between March and June 2005. This is consistent with the change found above by the SEC and by Diether et al. (2008, 2009). It should be noted that short selling increases almost another 1% between June and December, 2005. Since Asquith et al. (2005) point out that there is a long-term trend towards increasing short sales and since the increase between June and December is after the pilot study was underway, it is not clear whether the initial increase is due to the change in regulation or the long-term trend. More importantly, we are skeptical that the large difference in classification of short sales in Table 3 (e.g., almost a 30% difference in NYSE Lee–Ready classification results for pilot securities relative to non-pilot securities) is explained by these much smaller increases in short selling.

4. Additional problems with trade classification schemes: trade initiation, order type, and the implications of short sales

The result that short sales are predominately classified as buyer-initiated seems counter-intuitive. Since short sellers must, as mentioned above, first locate a security to borrow and accept a below-market rebate rate, a short sale would seem to be seller-initiated. Further, while it would seem reasonable to consider the first mover to be the “initiator,” the empirical literature that tests the Lee–Ready and other algorithms defines the initiating party as the one to last enter into the transaction. While there is no theoretical basis for this definition, the rationale is that the initiator is the trader who demands immediate execution and is willing to pay an “immediacy premium.”¹⁰ This definition is then used when testing the classification algorithms for accuracy, where accuracy is defined as the empirical results being consistent with an author’s inferential scheme.

The mechanics of the uptick and bid–price rules make this definition particularly problematic for short sales. If a short seller desires immediate execution of a sell order, it will not be filled if the prior trade was a downtick or zero-downtick. In this instance, the short seller must wait for an uptick before executing the trade. If, absent regulation, the short seller would have accepted the prevailing market price, they would have been the last party to enter into the transaction. Short sale restrictions, however, force the short seller to wait and effectively provide liquidity. When the trade is executed following the next uptick, the short seller may not be the last party to enter into the transaction. In this way, the

⁹Since most of our stocks have different floor traders, we also calculate classification rates by stock. Although these rates are not reported in a table, examining trade classification rates cross-sectionally reveals that there is no significant disparity between traders before and after regulation SHO took effect.

¹⁰Both Lee and Radhakrishna (2000), Odders-White (2000), and Ellis et al. (2000) consider the last party to enter a trade to be the initiator. Odders-White (2000) explicitly makes use of the concept of an “immediacy premium,” and Lee and Radhakrishna specify that Lee–Ready is useful in identifying “the more aggressive side of a trade.”

uptick and bid–price rules interfere with the mechanism that trade-classification algorithms use to identify the initiating party.

Even without the uptick and bid–price rules, there is still a problem when using the concept of an immediacy premium to classify short sales. All short sellers, both pilot and non-pilot, are required to find a security to borrow and then pay a premium in the form of a lower rebate rate on collateral. For a “non-special” stock, this premium is 10–15 basis points. For a “special” stock, this premium can be up to 700 basis points. Consequently, the premium paid by many short sellers is larger than that paid by the last trader for immediate execution, which is, at a maximum, the bid–ask spread. The fact that some short sellers may pay a larger cash premium than the last party to enter a transaction negates one of the underlying rationales upon which the trade classification algorithms depend. If “premium paid” is the defining determinant of initiation, then many short sales are indeed initiated by the seller even if they are not the last party to enter.

There is an additional problem involving trade classification that does not involve short sales. Identification of this problem is not new to the trade classification literature, but it is largely ignored despite its importance. Defining the last mover as the one who initiates a trade is most valid for single-sided market order transactions. These are trades that feature a specialist or market-maker that passively supplies liquidity on the other side. However, the ability of this rule to identify the initiating party is impaired for other order types. A market “cross,” for example, matches an existing buy and sell order. In such a situation, the buyer and seller effectively arrive at the same time, since both orders are already outstanding. Consequently, neither party initiated the trade. This type of ambiguity is why Lee and Radhakrishna (2000) and others limit their samples by removing market “crosses,” stopped trades, special orders, and multiple party market orders when testing the Lee–Ready algorithm. Lee and Radhakrishna (2000) eliminate 40% of all trades, while Finucane (2000) and Odders-White (2000) both eliminate 25%.

Recent authors do not limit their samples to single-sided market orders. This is because more recent trade databases, such as the TAQ database, do not include this information, while the TORQ database identifies order type. As a consequence, the application of trade classification algorithms to the TAQ data utilizes all trades, not merely those involving single-sided market orders. Therefore, an initiator is identified for a large number of trades where the algorithms are undefined. This inclusion means that the error rate for Lee–Ready and other algorithms will be higher for recent studies than that reported in the literature reviewed above, which predominantly used the TORQ data. This implication is troubling for all recent applications of trade algorithms.¹¹

This analysis and the results above have implications for researchers using the Lee–Ready and other algorithms to establish trade directions. First, the inclusion of non-market orders, i.e., those where there is more than one active trader, should cause the error rate to be higher than that reported previously. If the algorithm is 85% accurate (which the early literature shows) for the 75% of the trades that are market orders, and if classification is randomly assigned with 50% accuracy for those 25% that are not, the overall accuracy rate drops to 76.25% (or the error rate rises to 23.75%).

Second, if short sales are misclassified because of the uptick or bid–price rules, the accuracy of the algorithms is further decreased. If we assume that non-regulated pilot

¹¹Boehmer et al. (2007) demonstrate that misclassifications due to use of the Lee–Ready algorithm lead to a downwardly biased estimate of the PIN variable.

security short sales are correctly classified, then the percentage of non-pilot short sales classified as buyer-initiated is incorrect. Table 3 shows that an additional 29.2% of NYSE non-pilot short sales are classified as buyer-initiated by Lee–Ready when compared with pilot short sales in June and December. If there are no pilot securities (which there weren't before 2005) and if the percentage of short trades is 30% (Table 1 shows they were 29.7% in 2005), 6.6% of all trades will be incorrectly classified as buyer-initiated by the Lee–Ready algorithm due to regulation. This is after eliminating the 25% of trades, discussed in the paragraph above, which are not market orders.

To determine the extent to which short sales affect the overall error rate of trade classification algorithms, it is necessary to estimate an error rate for long trades. It is not a direct calculation but can be extrapolated from the studies using the 1991 TORQ database. As mentioned, those studies found a 15% error rate, at best, for the 75% of trades evaluated. The 15% is the sum of value-weighting the error rates on short trades and long trades. Assuming that the error rate of short sales caused by regulation is 29.2%, the error rate on long trades would be 13.89% or approximately half that of short sales.¹²

Applying this analysis to all trades in 2005 results in a much higher error ratio than the 15% reported in the literature. Assuming that 25% of trades are not classifiable (i.e., market crosses or stopped trades), 30% of all trades are short sales, 29.2% of short sales are misclassified due to regulation, and 13.8% remains the error rate on long trades, then the error rate for all trades is 26.3%. Analyzing this 26.3% rate, we find that approximately 12.5% is due to applying the algorithms to non-market order trades, 7.2% is due to errors in classifying long trades, and 6.6% is due to errors in classifying short sales.

In this estimation, we have made no attempt to address the problem of those short sales improperly classified as buyer-initiated even though short sellers paid a higher premium than the last party to enter a transaction. Doing so would increase the error rate. Thus, if it is true that the Lee–Ready algorithm, when applied exclusively to market orders in the TORQ database, is 15.0% inaccurate, applying the same algorithm to all trades in the TAQ database and accounting for the rise in short sale volume will increase the error rate by more than 75% to 26.3%.

We should note here that the SEC's decision to eliminate the uptick rule on July 6, 2007 does not reduce the importance of this conclusion to the empirical market microstructure literature. The vast majority of papers testing or applying the Lee–Ready and other algorithms have used and will continue, for the foreseeable future, to use data from periods where the uptick rule was in effect. As a result, the need to reevaluate the conclusions in these papers remains. It should also be noted that all three trade classification algorithms label the majority of pilot short sales as buyer-initiated. Thus, even without the uptick and bid–price rules, there appears to be a misclassification of short sales.

Further demonstrating the usefulness of the critiques of trade classification algorithms in this paper is the fact that a number of foreign markets still have price restrictions on short sales. Major stock markets in Australia, Japan, Canada, and other countries have rules

¹²This assumes a division of long trades to short trades of 92.1% versus 7.9%. While there is no time series of short sale percentages before 2005, the time series of NYSE short interest, should be correlated. Asquith et al. (2006) show that the average short interest in 2005 is 3.4%, while Asquith et al. (2005) show it was 0.9% in 1991. This implies a level of short sales in 1991 of 7.9%. Assuming an error in classifying short sales due to regulation of 29.2%, this means short sales contribute 2.3% of the 15% error rate and long trades 12.7%. This implies an error rate on long trades of 13.8%.

that are effectively the equivalent of the uptick rule. In addition, recent deterioration in U.S. financial markets has prompted calls for the reinstatement of the uptick rule. The effect that any new restriction on short selling could have on trade direction algorithms should be evaluated in light of the considerations raised in this paper. Finally, there remains the problem of applying the trade classification algorithms to all trades, which is usually done with TAQ data, regardless of current trading regulations.

5. Conclusion

In this paper, we test the Lee–Ready algorithm, the quote test, and the tick test on a sample of short sales. We find that all three methods of assigning trade direction classify a majority of short sales as buyer-initiated and that in many instances, these trades may be misclassified. For a random sample of 200 stocks trading over a three-month period in 2005, the Lee–Ready algorithm, with no delay, classifies 33.4% of our short sales as seller-initiated, the highest of any algorithm. For those short sales that are exempt from the uptick and bid–price rules under Regulation SHO, we see an increase in the percentage classified as seller-initiated. We find that, for NYSE securities, a short sale is more than three times as likely to be classified as seller-initiated when the security is exempt from the uptick rule (11.8% and 40.6% for June and 14.6% and 44.2% for December). Though the difference is not as striking for NASDAQ securities, the effect is significant and in the same direction.

In addition, there are problems when applying the trade direction algorithms to all trades in a dataset. Because the definition of “initiation” is predicated on the order in which the parties enter into a transaction, certain order types cannot be classified as either buyer- or seller-initiated. For example, referring to a market cross or stopped trade as either buyer- or seller-initiated is meaningless. Since most commercial datasets (TAQ being the most widely used) do not distinguish between order types, papers using Lee–Ready and other algorithms cannot exclude these order types from their samples.

Assuming a true error rate of 15% for trade classification, and considering only the effects of misclassification due either to the uptick and bid–price rules or the inclusion of all trades, we obtain a lower bound error rate of 26.3%. This percentage is considerably higher than the error rates found in previous empirical tests of the Lee–Ready and other trade classification algorithms. Further, the literature sometimes relies upon the definition of an initiator as the party that pays a premium for execution. Since it is possible that a short seller’s premium (derived from locate and borrowing constraints) may sometimes exceed the “immediacy premium,” some additional short sales may be incorrectly classified as buyer-initiated.

These results have potential implications for a large body of academic research. The Lee–Ready algorithm is widely used in the empirical market microstructure literature and, most recently, in the probability of informed trading (PIN) literature. Our results show that this research is subject to systematic bias in the treatment of short sales. The degree of this bias will vary and depends largely on the focus of the research and its dependence on correct assignment of trade direction. The effect of this inaccuracy, however, is not limited to published and working papers that rely on these algorithms. It is likely that misclassifications have also had an effect on research, which has been abandoned as a result of the insignificant results obtained through these algorithms.

Moreover, the extent of this problem has most likely become worse over time. Short sales have been rising as a percentage of market volume.¹³ As short transactions continue to constitute larger percentages of market volume, the trade direction algorithms will become increasingly inaccurate. Additionally, the results before 2005 would have been far worse since there were no pilot securities. Exchanges have also decimalized stock prices since the last domestic test of the Lee–Ready algorithm. Decimalization has narrowed quote spreads, and empiricists have typically found the trade direction algorithms to be less accurate between the bid and ask. For all of these reasons, it is likely that the performance of the trade algorithms has worsened over time.

Lee–Ready and other trade classification algorithms are essential to empirical research in several bodies of literature. This paper highlights several problems related to their application, particularly to short sales. It is necessary for researchers to recognize these problems when interpreting their empirical results.

Acknowledgments

We would like to thank Thomas Covert, Amy Edwards, SP Kothari, Ryan LaFond, Parag Pathak, Mark Ready, Jeri Seidman, the editor Avanidhar Subrahmanyam, Erik Theissen, and Jiang Wang for helpful discussions and comments. We would also like to thank Ted Keith for research assistance.

References

- Aitken, M., Frino, A., 1996. The accuracy of the tick test: evidence from the Australian Stock Exchange. *Journal of Banking and Finance* 20, 1715–1729.
- Alexander, G., Peterson, M., 2008. The effect of price tests on trader behavior and market quality: an analysis of Reg SHO. *Journal of Financial Markets* 11, 84–111.
- Asquith, P., Au, A., Pathak, P., 2006. Are short sale constraints binding?. MIT working paper.
- Asquith, P., Pathak, P., Ritter, J., 2005. Short interest, institutional ownership, and stock returns. *Journal of Financial Economics* 78, 243–276.
- Bessembinder, H., 2003. Issues in assessing trade execution costs. *Journal of Financial Markets* 6, 233–257.
- Bessembinder, H., 2003. Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis* 38, 747–777.
- Boehmer, E., Jones, C., Zhang, X., 2008. Which shorts are informed?. *Journal of Finance* 63, 491–527.
- Boehmer, E., Grammig, J., Theissen, E., 2007. Estimating the probability of informed trading—does trade misclassification matter?. *Journal of Financial Markets* 10, 26–47.
- Diether, K., Lee, K., Werner, I., 2008. Short-sale strategies and return predictability. *Review of Financial Studies*, 22, 575–607.
- Diether, K., Lee, K., Werner, I., 2009. It's SHO time! Short-sale price-tests and market quality. *Journal of Finance*, 64, 37–73.
- Ellis, K., Michaely, R., O'Hara, M., 2000. The accuracy of trade classification rules: evidence from Nasdaq. *Journal of Financial and Quantitative Analysis* 35, 529–551.
- Finucane, T., 2000. A direct test of methods for inferring trade direction from intra-day data. *Journal of Financial and Quantitative Analysis* 35, 553–576.

¹³Asquith et al. (2005) show that short interest rose dramatically from 1980 through 2003. Boehmer et al. (2008) report that short sales are 12.9% of NYSE volume during the period from January, 2000 through April, 2004, but that the percentage for the first four months of 2004 is 17.5%. They comment that short sales become “more prevalent as the sample period progresses.” As mentioned above, Asquith et al. (2006) and Diether et al. (2008) show that short sales constitute approximately 28% of all share trade volume in 2005.

- Lee, C., Radhakrishna, B., 2000. Inferring investor behavior: evidence from TORQ data. *Journal of Financial Markets* 3, 83–111.
- Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. *The Journal of Finance* 46, 733–746.
- Odders-White, E., 2000. On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets* 3, 259–286.
- SEC Office of Economic Analysis, 2007. Economic analysis of the short sale price restrictions under the Regulation SHO pilot.
- Theissen, E., 2001. A test of the accuracy of the Lee/Ready trade classification algorithm. *Journal of International Financial Markets, Institutions & Money* 11, 147–165.
- Vergote, O., 2005. How to match trades and quotes for NYSE stocks. Working Paper, Catholic University of Leuven.