ELSEVIER

Journal of Financial Markets 3 (2000) 259–286

Journal of FINANCIAL MARKETS

www.elsevier.nl/locate/econbase

On the occurrence and consequences of inaccurate trade classification☆

Elizabeth R. Odders-White*

Department of Finance, University of Wisconsin, Madison, 975 University Avenue, Madison, WI 53706, USA

Received 1 November 1995; accepted 1 March 2000

Abstract

The validity of many economic studies hinges on the ability to properly classify trades as buyer or seller-initiated. This study uses the TORQ data to investigate the performance of the Lee and Ready (1991, Journal of Finance 46, 733–746.) trade classification algorithm. I find that the algorithm correctly classifies 85% of the transactions in my sample, but systematically misclassifies transactions at the midpoint of the bid–ask spread, small transactions, and transactions in large or frequently traded stocks. I then provide evidence of the biases induced by inaccurate trade classification. © 2000 Elsevier Science B.V. All rights reserved.

JEL classification: G10

Keywords: Lee and Ready algorithms; Buyer/seller initiated trades

The validity of many economic studies hinges on the ability to accurately classify trades as buyer or seller-initiated. The importance of accurate trade

☆This work is part of my dissertation. I gratefully acknowledge the guidance and input of my committee members Mitchell Petersen, Kathleen Hagerty, and Leonard Soffer, and especially my chairpeople Laurie Simon Hodrick and Robert Korajczyk. I am also deeply indebted to Kenneth Kavajecz for providing data and for many invaluable comments. I appreciate helpful suggestions from Torben Andersen, Matthew Clayton, Joel Hasbrouck, Bjorn Jorgensen, Jennifer Lynch Koski, Mark Ready, Avanidhar Subrahmanyam (the editor), and an anonymous referee. All remaining errors are, of course, my own. This work was supported by an AAUW Educational Fund American Fellowship.

* Corresponding author. Tel.: + 1-608-263-1254; fax: + 1-608-265-4195.

E-mail address: ewhite@bus.wisc.edu (E.R. Odders-White).

PII: S1386-4181(00)00006-9

classification to market microstructure research is clear, but the significance extends beyond traditional microstructure studies.[1] Despite the importance of trade classification to economic research, the available data do not generally contain this information. Lee and Ready (1991) examined a pair of commonly used algorithms, namely the quote method and the tick method, which classify transactions based on execution prices and quotes. Lee and Ready then recommended that a combination of the two algorithms be used in practice (hereafter referred to as the Lee and Ready method).

The widespread use of these trade classification algorithms warrants an evaluation of their performance, with a focus on the effects of inaccurate trade classification on the results of existing studies. Specifically, the inclusion of misclassified transactions in a data set can cause one of two different types of problems: noise or bias. If the probability of misclassification is the same for all types of trades (e.g. large buys occurring the in the morning are as likely to be misclassified as small sells occurring in the afternoon), then trade misclassification will simply add random error to the data. If instead, particular types of transactions are more likely than others to be misclassified, then trade misclassification will add systematic error to the data and may ultimately bias the results.

Using the TORQ (Trades, Orders, Reports, and Quotes) database from the NYSE, which makes the direct determination of the initiator of a transaction possible, I evaluate the overall performance of the Lee and Ready algorithms and examine the consequences of misclassification. I find that the quote method misclassifies 9.1% of the transactions in my sample and fails to classify 15.9% of the transactions. The tick method misclassifies 21.4% of the transactions, and the combination recommended by Lee and Ready misclassifies 15.0%. Moreover, transactions inside the bid–ask spread, small transactions, and transactions in large or frequently traded stocks are especially problematic. I also provide evidence that misclassification can bias results. Specifically, I demonstrate that the increased buying found by Lee (1992) surrounding bad earnings announcements is due at least in part to small sales being misclassified as purchases by the Lee and Ready algorithm. In addition, in the context of a two-component Glosten and Harris (1988) model, I show that use of the Lee and Ready method results in a one-cent overestimation of the transitory component of the spread.

In a contemporaneous study, Lee and Radhakrishna (1996) conduct an evaluation of the Lee and Ready method for the NYSE and find that 93% of the

[1] Existing microstructure studies that utilize trade classification include Glosten and Harris (1988) and Hasbrouck (1988), among many others. Uses outside of typical microstructure settings include examinations of trading activity surrounding earnings announcements (Lee, 1992), studies of the effects of index futures on the underlying stock market (Choi and Subrahmanyam, 1994), and investigations of the aftermarket support of initial public offerings (Schultz and Zaman, 1994).

transactions in their sample are correctly classified. While the results of their study are quite informative, the focus of their paper differs somewhat from mine. First, they do not provide a detailed breakdown of the types of transactions that are misclassified. Second, although they also use the TORQ database, they focus on a smaller subset of the data. I present evidence in Section 4 that the trades eliminated by Lee and Radhakrishna tend to be misclassified more frequently. Finally, Lee and Radhakrishna do not investigate the consequences of trade misclassification.

The accuracy of trade classification algorithms has also been examined in other markets. For example, Aitken and Frino (1996) provided a detailed analysis of the performance of the tick method on Australian stock market data and found that it correctly classified 75% of the transactions in their sample. They also found that it was less accurate for small transactions and seller-initiated trades. In addition, Ellis et al. (2000) studied the accuracy of the quote, tick, and Lee and Ready methods on Nasdaq data. Overall, our results are strikingly similar. First, they documented accuracy rates of 78% for the quote method (when unclassified midpoint trades are labeled as misclassified), 80% for the tick method, and 83% for the Lee and Ready method. Second, they, too, found that trades that occur inside the spread or when trading is frequent are more likely to be misclassified.

The remainder of the paper is organized as follows. Section 1 provides a formal definition of the term 'initiator' as it is used here. Section 2 presents the Lee and Ready trade classification algorithms. The data and methodology are discussed in Section 3. Section 4 contains an analysis of the results and two sample applications illustrating the effects of misclassification. Section 5 concludes.

1. 'Initiator' defined

The goal of trade classification is to correctly determine the initiator of the transaction. Although the concept of a trade initiator is used throughout the finance literature, a formal definition of the term is rarely stated. No examination of the accuracy of trade classification algorithms can be conducted, however, without an explicit definition of the term 'initiator'.

One way to describe initiators is as traders who demand immediate execution (hereafter, the *immediacy* definition). A natural consequence of this definition is that traders placing market orders (or limit orders at the opposite quote) are labeled the initiators, and traders placing limit orders are viewed as non-initiators or passive suppliers of liquidity. A variant of this definition is used by Lee and Radhakrishna (1996) in their evaluation of the Lee and Ready algorithm.

Problems with this definition arise, however, when market orders cross, when limit orders are matched with other limit orders, and when market orders are

stopped, all of which can occur frequently. In the TORQ data, crossed market orders represent almost 12% of the transactions involving market and/or limit orders on both the buy and the sell sides, and limit orders matched with limit orders constitute another 17%. In addition, Ready (1999) found that 29% of the market orders in the TORQ data are stopped. Lee and Radhakrishna circumvent this problem by focusing only on transactions that take place between a 'clearly active' trader and a 'clearly passive' trader, thereby eliminating most of these transactions from their study. Unfortunately, studies that utilize trade classification algorithms apply the algorithms to all transactions, not to a select subset. Furthermore, if the data being used contain the order information necessary to distinguish between the active and the passive side, then trade classification algorithms are unnecessary. Consequently, in this paper I use the following, more general, definition of initiator:

Definition. The *initiator* of a transaction is the investor (buyer or seller) who placed his or her order last, chronologically.

The intuition behind the definition above (hereafter, the *chronological* definition) is very similar to that behind the immediacy definition. In both cases, the initiator is the person who caused the transaction to occur. In other words, by placing an order, the initiator determined the price and/or timing of the transaction. In fact, the two definitions are equivalent in many cases. For example, consider the transaction record in Fig. 1.

The buy limit order was placed at 1:02:55 and was matched with the standing sell limit order that had been placed approximately 2 h earlier. Consequently, this transaction is classified as buyer-initiated using the chronological definition above. The transaction in this example would also be classified as buyer-initiated using the immediacy definition, since the buy limit order was placed at the prevailing ask quote.

The advantage of the chronological definition is that it can be applied when the immediacy definition cannot. For example, when a market order is stopped and then executes against a subsequently placed limit order, the immediacy definition is unclear. Using the chronological definition, the placer of the limit order initiated the trade. This is consistent with the spirit of the immediacy

Ticker Symbol	Transaction Date	Transaction Time	Execution Price	Bid	Ask	Buy Order Date
MON	901114	1:03:10	5.75	5.50	5.75	901114
Buy Order Time	Buy Order Type	Buy Limit Price	Sell Order Date	Sell Order Time	Sell Order Type	Sell Limit Price
1:02:55	Limit	5.75	901114	11:07:55	Limit	5.75

Fig. 1. Sample transaction.

definition, since the investor who placed the market order is willing to wait for a chance at a better price.

2. The Lee and Ready algorithms

The contribution of the Lee and Ready study is twofold. First, they demonstrated that because updated quotes are often reported before the transactions that triggered them, a comparison of the execution price to the quotes in effect at the time of the transaction is inappropriate. This problem arose because quotes were updated on a computer inside the specialist's post, while transactions were recorded manually and fed into a reader alongside the specialist. The solution they proposed is the so-called '5-second rule', which directs that execution prices be compared to quotes reported a minimum of 5 s before the transaction was reported.

Second, Lee and Ready investigated two common methods for classifying trades, namely, the quote and tick methods. The quote method uses the following criteria to classify transactions: transactions above the spread midpoint, including those at the ask, are classified as buys; transactions below the spread midpoint, including those at the bid, are classified as sells; and transactions at the spread midpoint, which constitute 15.9% of the transactions in my sample, are left unclassified. All of the comparisons above employ the 5 second rule. Fig. 2 provides a graphical representation of the quote method.

Lee and Ready also investigated the tick method, which classifies transactions by comparing the price of the current trade to the price of the preceding trade. Upticks (price increases relative to the previous transaction price) are buys. Downticks (price decreases relative to the previous transaction price) are sells. Zero-upticks (zero price changes in which the last price change was an uptick) are buys and zero-downticks are sells. The advantages of the tick method are that it requires only transaction data (quotes are not necessary) and that

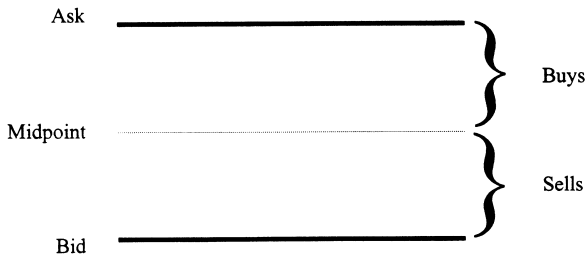

Fig. 2. The quote method.

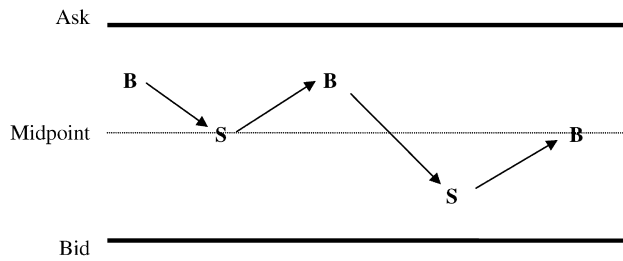

Fig. 3. The tick method.

no trades are left unclassified. The disadvantage is that the tick method incorporates less information than the quote method since it does not use the posted quotes. Fig. 3 contains a graphical representation of the tick method.

After carefully analyzing the quote and tick methods, Lee and Ready recommended that a combination of these algorithms be used in practice (the Lee and Ready method). Specifically, they suggested that the quote method be used to classify all transactions possible, and that the midpoint trades (left unclassified by the quote method) be classified using the tick method. Their recommendation was based on the following observations. First, they noted that 'the primary limitation of the tick test is its relative imprecision when compared to a quote-based approach'. This implies that the quote method should be employed whenever possible. Furthermore, in the context of a simple model, they demonstrated that the tick test correctly classified roughly 85% of trades occurring at the spread midpoint. The high predicted rate of accuracy of the tick method for midpoint trades, along with the likely superiority of the quote method, suggested that the proposed combination of the two was optimal.

The analysis conducted by Lee and Ready has proven extremely valuable to those conducting financial research – especially in the area of market microstructure – because it offers clear guidance regarding how to classify trades and how to properly align quote and transaction data. Prior to their analysis, researchers had little information on which to base such methodological decisions. Lee and Ready recognized that these algorithms were imperfect, however, and emphasized the difficulty in truly evaluating their performance without data on the true trade classification.

3. Data and methodology

The sample for this study comes from the TORQ database, which contains data on 144 NYSE stocks for the period from November 1, 1990 to January 31,

1991. The TORQ data consist of transaction, quote, and order records for all orders placed through one of the automated routing systems, as well as audit trail data, providing information on the parties involved and other detailed information about the trades.[2]

Before the Lee and Ready algorithms are evaluated, the true classification of each transaction is determined using a two-step process. In the first step, transaction records from the TORQ audit file are matched with order execution records from the TORQ order file, which contain the dates and times at which the executed orders were placed, as well as the order types (market, limit, or nonstandard). This information is then used in the second step to identify the initiator of each transaction according to the definition in Section 1, by comparing the order dates and times for the buy and sell sides of the transaction. For example, if the sell order is placed on November 1, and the buy order is placed on November 2, then the trade is buyer-initiated. Because the concept of an initiator is not applicable at the open (due to the opening auction), transactions occurring during the first 15 min of trading are excluded from the analysis.

Some transaction records cannot be matched with order execution records because at least one of the orders was not placed through an automated routing system. (For example, the order(s) may have been placed by a floor broker.) As a result, corresponding order execution records do not exist, and order information is not available for these transactions. Table 1 contains a breakdown of the magnitude of this problem by firm size. Panel A presents both the number and the percentage of transaction records for which the true initiator cannot be determined. Overall, the true initiator is unknown for 25.1% of the transactions. Panel B contains the number and percentage of buy and sell order execution records that remain unmatched, again broken down by firm size. Note that for the entire sample, there are only 4802 unmatched order execution records (2505 buys and 2297 sells), while there are 106,413 unclassified transaction records. This confirms that the true initiator cannot be determined for these transaction records primarily due to the lack of corresponding order execution records. Without the order date and time, the true initiator of the transaction cannot be determined. Transactions for which neither the buy nor the sell quantities were compared (agreed upon by both parties) also remain unclassified. Such transactions account for 4.1 of the 25.1%.

In the final step, the quote and tick algorithms, as well as the Lee and Ready algorithm, are applied to the transaction data to obtain the estimated classifications. Transactions for which the initiator cannot be determined are eliminated from the sample *after* the trade classification algorithms are applied. The resulting classifications are then compared to the true classification for each trade.

[2] For a description of the TORQ database, see Hasbrouck (1992).

Table 1
Determination of true classification

The true classification of each trade is determined by matching order execution records to transaction records. Some records cannot be matched because at least one of the orders was not placed through an automated routing system. The table below describes the magnitude of this problem. Panel A presents the number and percentage of transaction records for which the true initiator cannot be determined, broken down into deciles by firm size. Panel B contains the number and percentage of buy and sell order execution records that remain unmatched, again broken down by firm size.

Panel A: Unclassified transactions

Firm size decile	Total transactions	Number (%) unclassified
All	424,777	106,413 (25.1)
1 (largest)	237,289	65,087 (27.4)
2	55,638	11,243 (20.2)
3	36,949	9359 (25.3)
4	25,089	6317 (25.2)
5	25,343	5859 (23.1)
6	18,119	3134 (17.3)
7	12,135	2154 (17.8)
8	4727	1032 (21.8)
9	5853	942 (16.1)
10 (smallest)	3635	1286 (35.4)

Panel B: Unmatched order execution records

Firm size decile	Total records		Number (%) unmatched	
	Buys	Sells	Buys	Sells
All	435,338	419,837	2505 (0.6)	2297 (0.5)
1 (largest)	245,571	244,178	1969 (0.8)	1574 (0.6)
2	60,831	51,873	140 (0.2)	251 (0.5)
3	33,868	34,197	142 (0.4)	115 (0.3)
4	24,358	22,962	52 (0.2)	75 (0.3)
5	25,901	25,581	71 (0.3)	86 (0.3)
6	18,397	16,943	25 (0.1)	44 (0.3)
7	12,964	11,451	33 (0.3)	37 (0.3)
8	4310	4204	9 (0.2)	23 (0.5)
9	6121	5625	10 (0.2)	7 (0.1)
10 (smallest)	3017	2823	54 (1.8)	85 (3.0)

4. Results

4.1. Occurrence of misclassification

Table 2 contains a comparison of the true classification (buy or sell) with the classification from each of the three algorithms. Based purely on the percentage

Table 2
Performance of the algorithms

The table below contains a comparison of the true classification (buy or sell) to the classification from the quote (Panel A), the tick (Panel B), and the Lee and Ready algorithms (Panel C). A description of these methods is contained in Section 2 of the text. Each entry contains the number and percentage of transactions in the sample that fall into the respective category. Analyses are based only on transactions for which the true initiator can be determined.

Method and classification	True buy		True sell	
	Number	Percent	Number	Percent
Panel A: Quote method vs. true classification				
Quote method: Buy	127,827	40.15	14,997	4.71
Quote method: Sell	13,893	4.36	110,870	34.82
Quote method: Unclassified	26,308	8.26	24,469	7.69
Panel B: Tick method vs. true classification				
Tick method: Buy	134,649	42.29	34,662	10.89
Tick method: Sell	33,379	10.48	115,674	36.33
Panel C: Lee and Ready method vs. true classification				
Lee and Ready method: Buy	144,348	45.34	24,183	7.60
Lee and Ready method: Sell	23,680	7.44	126,153	39.63

of transactions classified correctly, the Lee and Ready method (Panel C) is the most accurate. The quote method (Panel A) performs relatively well on the transactions that it classifies, misclassifying only 9.1% of the transactions in the sample (4.36% from buys plus 4.71% from sells $\cong 9.1\%$). The quote method leaves almost 16% of the transactions unclassified, however. The tick method (Panel B) misclassifies $10.48\% + 10.89\% \cong 21.4\%$ of the transactions, while the Lee and Ready method misclassifies only $7.44\% + 7.60\% \cong 15.0\%$ of the transactions. Note that the percentage of misclassified transactions is fairly symmetric, with sells being misclassified as buys slightly more often than the reverse by all three methods. Since the Lee and Ready method is the most accurate method overall and is used most often, the remainder of the discussion focuses on this algorithm.

Recall that Lee and Radhakrishna (1996) found a 93% accuracy rate for the Lee and Ready method. Their accuracy rate exceeds the 85% rate found here because the trades that they eliminate are more likely to be misclassified by the algorithm. Specifically, the subset that they eliminate has an 81.5% accuracy rate, as opposed to the 93% rate they documented using their subsample. This difference is not driven entirely by the inclusion of stopped orders in my sample.

Despite its relatively good performance, the Lee and Ready method misclassifies 15% of the transactions in my sample, which amounts to almost 50,000 incorrectly labeled transactions. The percentage of transactions classified correctly is not the only measure of accuracy, however. For example, if the 50,000 transactions misclassified by the Lee and Ready method constitute a representative cross-section of the entire sample, then the misclassification will simply add noise to the data. In this case, the 85% accuracy rate is quite good. If, on the other hand, the Lee and Ready method systematically misclassifies certain types of transactions, a bias could result. In particular, if 'crucial' data points are frequently misclassified, then its 85% accuracy rate is not at all indicative of its true performance. Consequently, stating that the Lee and Ready method performs well could be misleading in the context of a specific application. Further investigation into the types of transactions that are misclassified is required to understand the degree of the bias induced by misclassification (if any). In particular, the importance of these transactions to different studies must be considered.

There are many dimensions along which transactions can be categorized to examine the accuracy of the algorithms. For example, the fact that the Lee and Ready method is based primarily on a comparison of execution prices to posted quotes suggests that trades at the quoted prices may be classified more accurately than those inside the spread. To test this hypothesis, I divide the sample into three groups: transactions that occurred at or outside the quotes, transactions that occurred at the spread midpoint, and transactions that occurred elsewhere inside the spread (not at the midpoint). Note that 0.6% of the 318,364 transactions in my final sample occur outside the posted quotes and that 96.8% of these result from order sizes that exceed the quoted depth.

Table 3 presents the frequency of misclassification for each of the subsamples. Transactions inside the spread are indeed misclassified more often than those at the quotes, and transactions at the spread midpoint are misclassified even more frequently. Of the 50,777 transactions with execution prices at the spread midpoint, 37.4% are incorrectly classified by the Lee and Ready method (as opposed to only 10.4% for transactions at the posted quotes). Interestingly, although the tick method correctly classified almost 80% of the transactions in the entire sample, it does not perform well when trades occur at the spread midpoint. These are exactly the transactions for which this method is being used in the Lee and Ready method. The poor performance of the algorithm for midpoint trades suggests that, under these circumstances, comparing the current transaction price to the previous price may be inappropriate.

Existing research suggests that small transactions, transactions in frequently traded securities, and transactions in large stocks may also be misclassified more often than others. Petersen and Fialkowski (1994) found that smaller trades were granted greater price improvement than larger trades. This is likely to result in a larger fraction of small trades occurring inside the bid–ask spread.

Table 3
Breakdown by transaction price in relation to quotes

The table below contains a breakdown of the accuracy of the Lee and Ready algorithm by price (relative to the posted quotes). Each row presents the number and percentage of transactions *in that category* that were correctly and incorrectly classified. Summing along each row provides the total number of transactions falling into the respective category. (Percentages sum to 100% along each row.) Summing down a 'Number' column yields the total number of correctly classified and incorrectly classified transactions in the sample. Analyses are based only on transactions for which the true initiator can be determined. The Chi-square statistic tests the hypothesis that the frequency of misclassification is independent of price.

Sample	Correct		Incorrect	
	Number	Percent	Number	Percent
At or outside the quotes	231,308	89.60	26,834	10.40
Inside the spread but non-midpoint	7389	78.23	2056	21.77
At the spread midpoint	31,804	62.63	18,973	37.37
	$\chi^2_2 = 24{,}507.472$		p-value = 0.001	

Since the algorithms are best suited for transactions at the quotes, small transactions may be misclassified more often than larger transactions.

Frequent trading may lead to higher misclassification rates for several reasons. First, if the 5 second rule is not appropriate, its use may induce misclassification by misaligning the quotes and trades. Second, the presence of a more active crowd on the trading floor for frequently-traded stocks may mean that more trades occur inside the spread, leading to higher misclassification rates. Finally, the rapidly changing quotes that stem from frequent trading may be problematic since the Lee and Ready algorithm often uses the quotes as a reference point.

Firm size is often viewed as a proxy for asymmetric information. If larger firms pose greater (lesser) adverse selection risks to market makers, then price improvement may be less (more) likely to occur. Consequently, transactions in large stocks would take place at the posted quotes more (less) often and would be misclassified less (more) frequently as a result. In addition, the strong correlation between firm size and overall trading activity suggests that large stocks may be misclassified more often.

To test these hypotheses, I divide my sample along four dimensions: trade size, time between trades, number of transactions during the sample period, and firm size.[3] Table 4 contains the results. The statistics in Panel A demonstrate

[3] Transactions were also broken down by the day of the week and time of day to examine whether any relation exists between interday and intraday trading patterns and misclassification. No significant patterns emerged across days of the week, but transactions in the morning were misclassified more frequently than those later in the day (16.31% at or before noon vs. 14.39% after noon), due in part to higher activity in the morning.

Table 4
Breakdown of Lee and Ready method misclassification by characteristics

The table below contains a breakdown of the accuracy of the Lee and Ready algorithm by the firm and trade characteristics described in Section 4 of the text. Each row presents the number and percentage of transactions *in that category* that were correctly and incorrectly classified. Summing along each row provides the total number of transactions falling into the respective category. (Percentages sum to 100% along each row.) Summing down a 'Number' column within a category (e.g. trade size) yields the total number of correctly classified and incorrectly classified transactions in the sample. Analyses are based only on transactions for which the true initiator can be determined. Chi-square statistics test the hypothesis that the frequency of misclassification is independent of the characteristic. The chi-square statistics are presented in the final column of the table and *p*-values are in parentheses.

Sample	Category	Correct		Incorrect		Chi-Square Stat (*p*-value)
		Number	Percent	Number	Percent	
Panel A: Trade size (share-based measure)						
Full sample	300 shares or less	114,651	83.15	23,232	16.85	$\chi^2_1 = 627.255$
	301 shares or more	155,850	86.35	24,631	13.65	(0.001)
Inside the spread only	300 shares or less	20,244	66.10	10,384	33.90	$\chi^2_1 = 28.283$
	301 shares or more	18,949	64.03	10,645	35.97	(0.001)
At or outside quotes only	300 shares or less	94,407	88.02	12,848	11.98	$\chi^2_1 = 494.205$
	301 shares or more	136,901	90.73	13,986	9.27	(0.001)
Panel B: Trading frequency – time between transactions						
Full sample	5 s or less	29,811	79.93	7,487	20.07	$\chi^2_2 = 970.286$
	6–30 s	56,852	84.27	10,616	15.73	(0.001)
	Over 30 s	183,838	86.07	29,760	13.93	
Panel C: Trading frequency – number of transactions						
Full	3000 or fewer	75,321	89.98	8,389	10.02	$\chi^2_2 = 3352.51$
Sample	3001–15,000	82,205	86.16	13,206	13.84	(0.001)
	Over 15,000	112,975	81.14	26,268	18.86	

Panel D: Time since quote update						
Full sample	1 min or less	73,127	82.52	15,486	17.48	$\chi^2_2 = 824.195$
	1–5 min	88,532	84.62	16,096	15.38	(0.001)
	Over 5 min	108,842	86.99	16,281	13.01	
One minute or Less Only	10 s or less	11,441	81.51	2,596	18.49	$\chi^2_2 = 21.591$
	10–30 s	29,990	82.27	6,461	17.73	(0.001)
	30 s or more	31,696	83.14	6,429	16.86	
Panel E: Firm size						
Full sample	Large (Deciles 1–5)	237,034	83.92	45,409	16.08	$\chi^2_1 = 2132.56$
	Small (Deciles 6–10)	33,467	93.17	2,454	6.83	(0.001)
Inside the spread only	Large (Deciles 1–5)	35,976	64.58	19,729	35.42	$\chi^2_1 = 80.981$
	Small (Deciles 6–10)	3,217	71.22	1,300	28.78	(0.001)
At or Outside quotes Only	Large (Deciles 1–5)	201,058	88.67	25,680	11.33	$\chi^2_1 = 1733.59$
	Small (Deciles 6–10)	30,250	96.33	1,154	3.67	(0.001)
3,000 Trades or less only	Large (Deciles 1–5)	44,670	87.94	6,126	12.06	$\chi^2_1 = 595.356$
	Small (Deciles 6–10)	30,651	93.12	2,263	6.88	(0.001)
3001–15000 Trades only	Large (Deciles 1–5)	79,389	85.92	13,015	14.08	$\chi^2_1 = 146.034$
	Small (Deciles 6–10)	2,816	93.65	191	6.35	(0.001)
Over 15,000 Trades only	Large (Deciles 1–5)	112,975	81.14	26,268	18.86	N/A
	Small (Deciles 6–10)	0	N/A	0	N/A	
Panel F: Accuracy by type of midpoint trade						
Midpoint trades only	Uptick or downtick	6734	76.89	2024	23.11	$\chi^2_1 = 918.924$ (0.001)
	Zero-tick (no price change)	25,070	59.66	16,949	40.34	

that small transactions are indeed misclassified more frequently than larger transactions. While 16.85% of transactions of 300 shares or less are misclassified by the Lee and Ready method, only 13.65% of transactions greater than 300 shares are misclassified. This is consistent with Aitken and Frino's (1996) results using the tick method on Australian data. The chi-square statistic in the final column tests for independence between the frequency of misclassification and trade size. The null hypothesis of independence is rejected at the 0.1% level.

As discussed above, small trades may be misclassified more frequently simply because they are more likely to occur inside the posted spread. Preliminary statistical evidence is consistent with this hypothesis. The probability of a 300-share or smaller trade occurring inside the spread is 22.2%, as opposed to 16.4% for larger trades. To test the hypothesis more directly, I partition the sample into transactions that occurred inside the spread and transactions that took place at (or outside) the quotes. Then, the relation between trade size and misclassification is examined within each subsample.

Panel A of Table 4 also contains the breakdown of misclassification by trade size for trades inside and at the quotes, respectively. The results in rows 5 and 6 demonstrate that small trades are more likely to be misclassified even when they occur at the quotes. In addition, conditional on the trade having occurred inside the spread, larger trades are actually *more* likely to be misclassified than smaller trades (rows 3 and 4). When non-midpoint transactions inside the spread are eliminated from the sample, however, I find no significant relation between trade size and misclassification. In aggregate, this evidence refutes the hypothesis that the association between misclassification and trade size is simply the result of small trades occurring inside the spread. In other words, the size of the trade affects the misclassification rate even after controlling for the price relative to the quotes.

The hypothesized relation between trading frequency and misclassification also exists. Two measures of trading frequency are used in this study: time between transactions and total number of transactions during the sample period. Panels B and C of Table 4 contain the results. Transactions occurring less than 5 s apart are misclassified 20.07% of the time, far more frequently than the other transactions. Firms with more transactions also have a greater incidence of misclassification in percentage terms (10.02% for firms with 3000 or fewer total transactions vs. 18.86% for those with over 15,000).[4]

The increased misclassification for trades occurring less than 5 s apart does not appear to be driven by a failure of the 5 second rule. In my sample, the 5 second rule takes effect for 13,156 of the 318,364 transactions (roughly 4%).

[4] I also investigated relative measures of trading activity (including the number of transactions in the given day divided by the daily average for the stock and the time between the current and prior transactions divided by the average time between transactions for the stock) but no significant patterns emerged.

The 5 second adjustment changes the Lee and Ready classification (relative to no adjustment) for only 1218 of those trades, however. Although 42% of the 1218 trades are misclassified, eliminating the 5 second rule induces more misclassification than it corrects. Furthermore, the possibility that a 10 second rule is more appropriate for small stocks (deciles 6–10) was investigated, but the 5 and 10 second rule classifications differed for only 42 (0.01%) of the transactions in the sample.

The more frequent misclassification during active trading is not due to a higher probability of occurring inside the spread, either. In fact, the probability of execution inside the spread is inversely related to both measures of trading frequency. For example, 15.76% of transactions occurring less than 5 s apart took place inside the spread, versus 20.30% of those more than 30 s apart. Similarly, for stocks with 3000 or fewer transactions, 20.5% occurred inside the spread, as opposed to 16.73% for stocks with over 15,000 transactions.

If the constant changing of quotes when trading is frequent is the source of the problem, then a direct relationship should exist between trading frequency and quote 'freshness' (how recently the quotes were updated), and also between quote freshness and misclassification. Quote freshness is, in fact, significantly positively associated with both measures of trading frequency. For example, for stocks with over 3000 transactions, only 29% of the trades take place more than 5 min after a quote revision, as opposed to over 68% for stock with 3000 or fewer transactions. Similarly, only 26.55% of transactions occurring within 30 s of the previous trade take place more than 5 min after a quote revision versus 45.56% when trades are over 30 s apart. In addition, Table 4 Panel D shows that the recent updating of quotes is associated with increased misclassification as well. In particular, 17.48% of transactions occurring 1 min or less after a quote change are misclassified versus only 13.01% of transactions with lag times over 5 min. The relation is also monotonic within the first category, with transactions taking place within 10 s of a quote change significantly more likely to be misclassified than the others. The relation between quote freshness and frequent trading does not completely drive the increased misclassification associated with frequent trading, however (results not shown).

Firm size also plays a role in misclassification (Table 4 Panel E), with large stocks misclassified much more frequently than small stocks (16.08% vs. 6.83%). Particularly striking is the fact that almost 95% of the transactions misclassified by the Lee and Ready method occur in stocks from the largest five deciles (45,409 of the 47,863 total). This is due only in part to the greater number of transactions for these stocks.

The increased misclassification of large stocks cannot be entirely explained by either a higher probability of occurring inside the spread or by the positive correlation between firm size and trading frequency (or by both). Although transactions in large stocks are more likely to take place inside the spread than small stocks (19.72 vs. 12.57% with a p-value of 0.001), the results in Panel E

demonstrate that this does not explain the relation between firm size and misclassification. After dividing the sample into transactions inside and at the quotes, the probability that large stocks' transactions are misclassified continues to exceed that of small stocks for both subsamples (35.42% vs. 28.78% inside the spread and 11.33% vs. 3.67% at the quotes). Similarly, although firm size and trading frequency are highly correlated, the statistics in Table 4 Panel E demonstrate that large stocks are misclassified more often even after controlling for trading frequency. Consequently, more frequent trading cannot fully explain the association between firm size and misclassification. (Likewise, firm size does not drive the increased misclassification associated with frequent trading.)

The results in Tables 3 and 4 demonstrate that the Lee and Ready method misclassifies transactions at the spread midpoint and transactions in large or actively traded stocks most frequently. In addition, small transactions at the quotes and, to a much lesser extent, large transactions inside the spread are problematic.[5] Some theoretical justification for these findings has been presented, but questions about the cause still remain.

It is clear that price improvement is at the root of the problem. Although we often think of price improvement as trades executing inside the spread, orders executing at the opposite quote (buying at the bid and selling at the ask) is simply a more extreme case of price improvement. In this sense, trades that are more likely to be misclassified are simply those that are more likely to receive some form of price improvement. For example, perhaps transactions in large firms are misclassified more often due to price improvement stemming from less uncertainty (e.g. more analysts following the stock) or more liquidity provision (e.g. floor traders).

One of the ways in which misclassification results from price improvement is examined in Fig. 4. The figure describes a common case of midpoint-trade misclassification. In this scenario, there is no change in the execution price and the spread widens (the bid decreases and the ask remains the same or the ask increases and the bid remains the same). In the example, the initial bid and ask

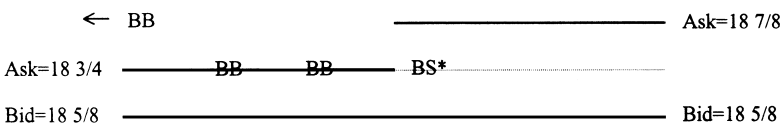

Fig. 4. Midpoint misclassification.

[5] When the analysis was repeated using only the Lee and Radhakrishna subsample, the results were equally as strong or stronger, with two exceptions. Using their subsample, time between transactions is no longer a statistically significant determinant of misclassification and large trades are misclassified slightly more frequently than small trades (7.2% vs. 5.7%).

are \$18 5/8 and \$18 3/4, respectively, and the ask increases to \$18 7/8. Transactions are indicated with two letters. The first letter represents the true classification (B = buy and S = sell). The second letter represents the Lee and Ready classification. The problematic transaction (which is labeled with a star in the diagram) is the buyer-initiated trade occurring immediately after the increase in the ask price. Since it is a midpoint transaction, it is classified using the tick method and is labeled as seller-initiated because it is a zero-downtick.

This systematic misclassification stems from the guidelines regarding specialist behavior at the NYSE. The specialist is required to keep a fair and orderly market, which includes posting reasonable depth. He can do so in two ways: by simply reflecting the limit order book (i.e. posting bid and ask prices equal to the best bid and ask on the limit order book and posting depths equal to the number of shares on the book at those prices) or by posting additional shares himself. Suppose the specialist chooses to reflect only the shares on the limit order book but feels that the shares at the best prices on the book do not provide sufficient depth. In this case, he may 'move up' the limit order book to post a price at which there are more shares on the book. In other words, he may choose to widen the spread by increasing the ask (and/or decreasing the bid) in order to post more depth. The result is one or more hidden limit orders on the side of the market in which the quote was moved. When these limit orders are hit, the transaction occurs at the spread midpoint. Transactions taking place under these circumstances are much more likely to occur within 30 s of a quote change (37.61%) than transactions in the full sample (15.86%).

Roughly, 10% of the misclassified transactions in my sample occur in this situation. Changes in the NYSE guidelines prohibiting specialists from hiding limit orders in this manner went into effect in 1996 and should improve the overall accuracy of the algorithm.

Also recall that the Lee and Ready algorithm employs the tick method for midpoint transactions and that this method classifies zero-tick trades (transactions for which there is no price change) by referring back to the last price change. Consequently, any midpoint transactions immediately following the trade in question (at the same price) will also be misclassified.

In fact, zero-tick trades are problematic in general because the prior trade is often an inappropriate benchmark. For example, if the prior trade took place long ago, it is 'stale' and does not reflect current market information. On the other hand, if trading is very active, situations like that in Fig. 4 may occur, in which two or more transactions pick off hidden limit orders at the midpoint. The results in Panel F of Table 4 verify that zero-tick midpoint trades are misclassified much more frequently than other midpoint trades (40% vs. 23%). In addition, these transactions account for 89% of the misclassification occurring at the spread midpoint. Fig. 4 is only one example of a trading pattern that induces misclassification. There are other such patterns, each accounting for a (sometimes small) fraction of the total misclassification.

The information contained in Tables 3 and 4 can be used by researchers to determine the best way to apply the Lee and Ready algorithm in future studies. In particular, I recommend that researchers partition their transaction samples along the dimensions investigated above and examine the impact on the results of their studies. If the findings are consistent across partitions, then researchers can be reasonably confident that their results are robust to misclassification bias. On the other hand, if the results change along these dimensions without any clear explanation given the focus of the research, this suggests that misclassification may be a problem. In this case, choices should clearly be guided by the goal of the study in question and the nature of the data. For example, eliminating midpoint transactions (effectively using the quote rule) is a good strategy in many cases. In situations where midpoint transactions are necessary, however, this is obviously not possible. At a minimum, differences across partitions should be discussed along with the overall results.

A few caveats are necessary at this point. First, the analysis above is based on a single data set (TORQ), which contains data for 144 stocks over a three-month period. Because these data are used to determine the true initiator of each transaction, I am implicitly assuming that the data accurately represent the truth. While no data set is error free, the TORQ data are quite clean and I have no reason to suspect that any non-random errors that could bias my results exist.

Another concern is the fact that only electronically-submitted orders are included in the data set and, as a result, not all transactions could be classified. Consequently, like Lee and Radhakrishna, I am unable to evaluate the performance of the algorithm for all transactions. To the extent to which the systematic misclassifications are actually idiosyncratic to my sample, the results will not generalize.

There is fairly substantial evidence to suggest that this is not the case, however. First, most of my results are consistent with existing theoretical and empirical evidence and stem from the market structure of the NYSE. Second, many of my findings are consistent with those of Aitken and Frino (1996) for the Australian Stock Exchange and Ellis et al. (2000) for Nasdaq. Finally, if the failure to classify all the trades in the sample creates any bias, it is against my results. In particular, unclassified trades tend to be in bigger firms (92.4% in deciles 1–5 vs. 88.72% for classified transactions), in more frequently traded stocks (33.2% less than 5 seconds vs. 11.72% for classified and 79.9% with over 3000 transactions vs. 73.71% for classified), at the spread midpoint (23.9% vs. 15.96%), and larger trades than those in the classified sample. This is not surprising since these types of orders are less likely to be submitted electronically. It is notable, however, because these are exactly the types of trades that tend to be misclassified by the algorithm. This suggests that the 15% misclassification rate is actually a conservative estimate.

In summary, while caution should always be exercised when drawing conclusions using a single sample, I am confident that my findings are not driven by any limitations of my data.

4.2. Consequences of misclassification

The preceding section contains detailed descriptions of the types of transactions for which the Lee and Ready method fails. Such an analysis seems unnecessary, however, if misclassification has no effect on the results of economic research. This section provides two applications in which misclassified transactions lead to biased results.

4.2.1. Investor behavior surrounding earnings announcements

The first example is Lee's (1992) study of investor behavior surrounding earnings announcements. Lee used event-study methodology to examine the intraday trading activity of large and small traders around both good and bad earnings announcements. As expected, he found that good earnings announcements were associated with periods of increased buying regardless of trade size. He also documented a puzzling increase in the number of small purchases in response to *bad* earnings announcements, however. Although he considered several possible explanations, he was ultimately unable to explain this result.

The analysis in Section 4.1 suggests that the more frequent misclassification of *small seller-initiated* transactions may be at least partially responsible for this anomaly. The intuition behind this explanation is as follows. First, the results in Table 3 demonstrate a greater incidence of misclassification for transactions occurring inside the spread. Second, we would expect *small* trades to occur inside the spread more often since smaller transactions tend to receive price improvement more often than large transactions. The results in Table 4 confirm this hypothesis. In addition, Petersen and Fialkowski found that *sell* orders received greater price improvement than buy orders. As a result, *small sell* transactions are likely to be misclassified (as *buys*) more often than other types of transactions.

In Appendix A, I provide two pieces of evidence in support of this hypothesis. First, I partition the data into large trades and small trades using Lee's dollar-based measure of trade size (which differs slightly from that used in Table 4). The results in Table 5 illustrate that the Lee and Ready method and the true classification produce almost identical fractions of buys and sells for large transactions. On the other hand, for small transactions, the discrepancy between the Lee and Ready and the true classifications is over three times as large, with the Lee and Ready method classifying more trades as sells. While the difference is small in absolute terms, it is big relative to that for large trades and is statistically significant at the 0.03% level. The frequency of misclassification is likely to increase around earnings announcements due to an increase in trading, as well.

Second, I replicate Lee's analysis to provide direct evidence that trade misclassification is a partial explanation of the anomaly documented in his study.

Table 5
Comparison of Lee and Ready and true classifications for large and small trades

Columns 2 and 4 (3 and 5) contain the number (percentage) of transactions in each subsample classified as purchases using the Lee and Ready method and the true classification, respectively. Numbers and percentages for sales are presented in an analogous fashion. The final columns test the hypothesis that the percentage of buys (or sells) using the Lee and Ready method is the same as the true percentage for the subsample.

Sample	Buys				Sells				H_0: LR% = True%	
	L&R		True		L&R		True		*T*-stat	*p*-value
	#	%	#	%	#	%	#	%		
Large trades	96,451	51.98	96,878	52.21	89,108	48.02	88,681	47.79	1.403	0.1606
Small trades	72,080	54.28	71,150	53.57	60,725	45.72	61,655	46.43	3.620	0.0003

Although the results are not statistically significant due to a limited sample size, I find that replacing the Lee and Ready classification with the true classification eliminates three of the five periods of abnormal buying activity for small trades surrounding bad announcements. While the evidence is not strong enough to conclude that Lee's result is driven entirely by inaccurate trade classification, it is certainly sufficient to document the impact of inaccurate trade classification on research.[6]

4.3. Components of the bid–ask spread

As discussed in the introduction, the Lee and Ready algorithm is likely to overestimate the costs associated with trading. This stems from the fact that the algorithm classifies trades that occur above (below) the spread midpoint as buys (sells). If a seller-initiated transaction occurs above the midpoint, it will be misclassified as buyer-initiated by the algorithm and price improvement will be underestimated (i.e. transaction costs will be overestimated).

I test the hypothesis that the use of Lee and Ready algorithm results in the overestimation of transaction costs by estimating a two-component model of the bid–ask spread (see Glosten and Harris, 1988) using both the Lee and Ready and the true classifications. The complete results are presented in Appendix B. I find that the use of the Lee and Ready method leads to a statistically significant one-cent overestimation of the order processing component of the spread on average. The adverse selection component is roughly the same using either classification. This implies that actual transaction costs are lower than those suggested by studies that utilize the algorithm. While \$0.01 is small in an absolute sense, it represents a significant difference in trading costs, particularly for large transactions.

5. Conclusion

The goal of this study is to examine the performance of trade classification algorithms and to determine the degree to which misclassification biases the results of economic research. I find that the trade classification algorithm recommended by Lee and Ready performs quite well in general, correctly classifying 85% of the transactions in my sample. The algorithm systematically misclassifies certain types of transactions, however. In particular, transactions inside the bid–ask spread, small transactions, and transactions in large or frequently traded stocks are often misclassified.

[6] Another partial explanation may be the sub-optimal individual investor behavior recently documented by Barber and Odean (2000), among others.

Evidence of the impact of inaccurate trade classification on economic research is provided. In a study of trading activity surrounding earnings announcements, Lee (1992) documented an anomalous finding that small trade volume consists predominantly of stock purchases, even when the announced earnings level is below expectations. I demonstrate that the increased buying found by Lee results at least in part from small sales being misclassified as purchases by the Lee and Ready algorithm. In addition, I document a one-cent overestimation of the order-processing component of the bid–ask spread using the Lee and Ready method.

In light of this evidence, I recommend that researchers partition their transaction samples along the dimensions investigated in the paper and examine the impact on the results of their studies. If the findings are consistent across partitions, then researchers can be reasonably confident that their results are robust to misclassification bias. On the other hand, if the results change along these dimensions without any clear explanation given the focus of the research, this suggests that misclassification may be a problem. In this case, at a minimum, differences across partitions should be discussed along with the overall results. In addition, authors could consider repeating their analyses after eliminating the subset of transactions that is most likely to be misclassified. Clearly, the optimal approach depends on the nature of the study, since omitting trades could introduce a much greater bias than it eliminates in some contexts.

Appendix A. Investor behavior surrounding earnings announcements

Lee (1992) divided transactions into small trades, a proxy for individuals, and large trades, a proxy for institutions, and studied the trading behavior of each group around good, bad, and neutral earnings announcements (relative to Value Line and other estimates). Surprisingly, he found that bad announcements are followed by an abnormally high number of small stock purchases. I propose trade misclassification as a partial explanation for this result.

I first investigate the hypothesis that small sales are systematically misclassified as buys by the Lee and Ready algorithm. Table 5 compares the Lee and Ready classification to the true classification for small and large trades. Lee defines the cutoff between small and large trades as the largest number of round lot shares that can be purchased for $10,000 or less, based on the stock price at the end of the sample period.

The results in Table 5 demonstrate that the Lee and Ready method classifies 427 fewer (more) transactions as buys (sells) than the true classification – a difference of 0.23% that is not statistically significant at any reasonable level. On the other hand, for small transactions, the discrepancy between the Lee and Ready and the true classifications is over three times as large, with the Lee and Ready

method classifying 930 (0.71%) more trades as sells. While this difference is still small in absolute terms, it is substantially larger and is statistically significant at the 0.03% level. The frequency of misclassification is likely to increase around earnings announcements due to an increase in trading, as well. This evidence suggests that misclassification could be driving Lee's finding that bad announcements are followed by abnormally high buying activity for small trades.

I investigate this hypothesis further by replicating Lee's study using the TORQ data. Lee computed mean abnormal directional imbalance (MAD) values for both small and large trades for each type of announcement as follows. He began by computing a trading direction measure (FDIR) for each half-hour interval surrounding the announcement (days -1 to $+3$) by subtracting the number of sell transactions from the number of buy transactions and scaling by the total number of trades of the given size (small or large) for the firm during the sample period. He then compared these measures to their averages over the non-announcement period ($\overline{\text{FDIR}}$) using a mean abnormal directional imbalance (MAD) metric. The mean abnormal directional imbalance is defined a

$$\text{MAD}_r^z = 1/m_r^z \sum_{i=1}^{m_r^z} (FDIR_{ik}^z - \overline{FDIR_{ik}^z}),$$

where r represents the announcement-interval (-13 to $+38$), i represents the firm, z represents the trade size (small or large), and k represents the time of day (13 intervals spanning the trading day). The denominator, m_r^z, is the number of announcements for which there was at least one trade of size z in announcement interval r.

The results I obtain when replicating Lee's study are contained in Figs. 5 and 6. Graphs for good and neutral announcements are excluded due to their similarity to the original Lee (1992) results. Fig. 5 focuses on large transactions. As expected, the graphs contained in Panels A and B (Lee and Ready and true classification, respectively) both document abnormal selling in the period following the earnings announcement.

Panel A of Fig. 6 displays the results for small transactions using the Lee and Ready classification. Here again, the results are similar to those found by Lee, with unexplained abnormal buying activity in the days following the announcement. Specifically, there are five post-announcement intervals with statistically significant buying activity. The results using the true classification differ substantially from the results using the Lee and Ready method, however (and, consequently, differ from Lee's findings). A comparison of Panel B with Panel A demonstrates that much of the abnormal buying activity disappears when the Lee and Ready classification is replaced with the true classification. In particular, the number of intervals with statistically significant positive MAD values decreases from five to two. None of the differences in MAD for the individual intervals is significant at conventional levels, however. I suspect that the lack of

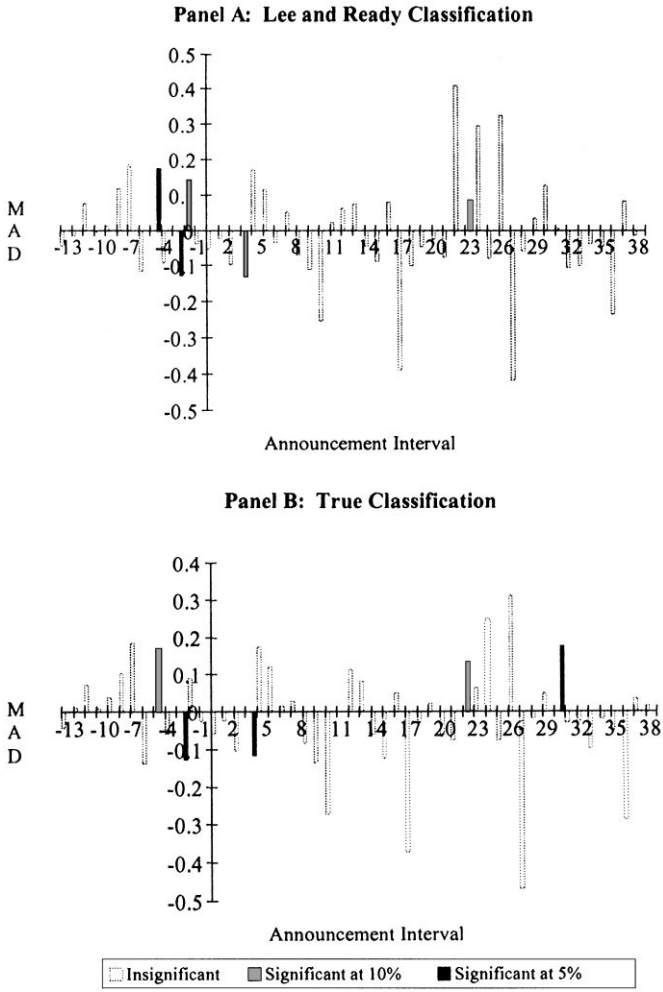

Fig. 5. Mean abnormal directional imbalance for large trades surrounding bad announcements.

statistical significance is due in part to the limited size of my sample. (I have only 23 bad earnings announcements in my sample, while Lee had 240 in his study.)

If one focuses exclusively on the MAD values that are significantly different from zero at the 10% level or less, then it becomes clear that the cumulative difference between the Lee and Ready values and the true values over the announcement period is positive for large transactions (0.092). This means that the Lee and Ready classification biases MAD values for large transactions downward (towards selling) overall. Conversely, for small transactions, the true

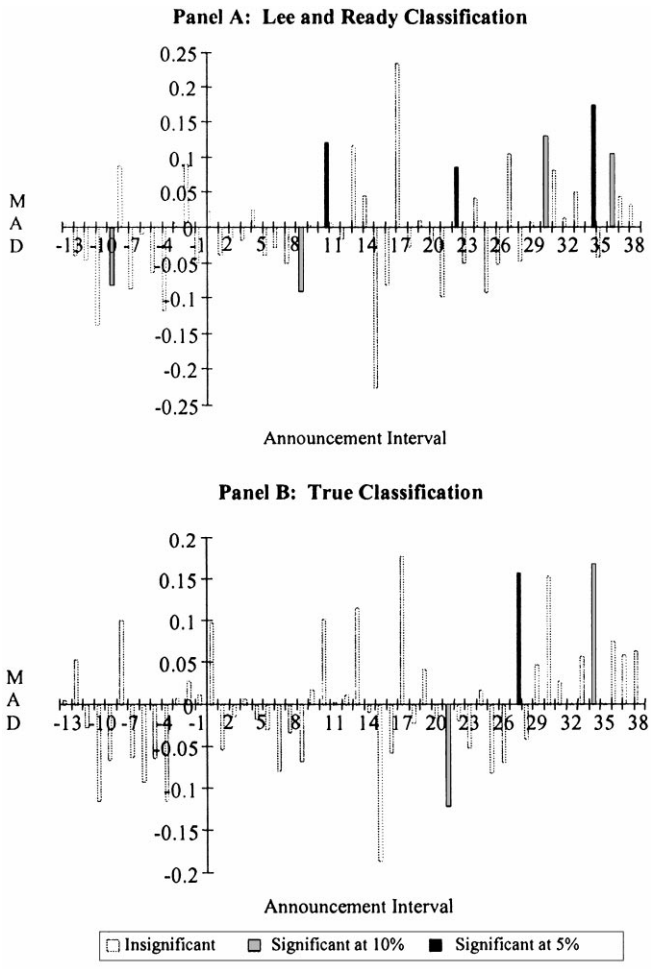

Fig. 6. Mean abnormal directional imbalance for small trades surrounding bad announcements.

MAD value is often smaller than the Lee and Ready value, with a cumulative difference of -0.238. In other words, the Lee and Ready classification biases MAD values for small transactions upward (towards buying). Furthermore, the magnitude of the bias for small transactions is 2.5 times larger than that for large transactions.

There is additional evidence lending support to trade misclassification as an explanation for the abnormal frequency of small purchases. Recall that the true initiator of the trade could not be determined for roughly 25% of the transactions in the sample. Consequently, one fourth of the transactions classified using

the 'true classification' were actually classified using the Lee and Ready method for the purposes of this application.[7] As a result, the two occurrences of abnormal buying may remain in Panel B of Fig. 6 due to biases from the Lee and Ready method. Although the limitations of the data make a perfect test of the role of misclassification in this anomaly impossible, the results provide strong evidence in favor of misclassification as a *partial* explanation.

Appendix B. Components of the bid–ask spread

Glosten and Harris (1988) estimated the following two-component model of the bid–ask spread:

$$\Delta P_t = c_0 \Delta Q_t + z_1 Q_t V_t + e_t,$$

where P_t represents the price for transaction t, $Q_t = 1$ for buyer-initiated transactions and -1 for seller-initiated transactions, V_t represents the size of transaction t in 1000 shares, c_0 is the transitory component of the spread, and z_1 is the adverse selection component of the spread.

The transitory component compensates the market maker for order processing costs, while the adverse selection component arises because market makers must trade with investors who are potentially better informed than they.[8]

I estimate the model for each of the 144 stocks in the TORQ data twice – first using the Lee and Ready classification to determine Q_t and then using the true classification. The results are contained in Table 6.

Panel A contains the mean transitory (order processing) and adverse selection component estimates across the 144 stocks in the sample, as well as the 25th, 50th, and 75th percentiles for those estimates. A comparison of the rows of the table shows that use of the Lee and Ready algorithm results in an estimated transitory component that is 1.14 cents too large on average. The adverse selection component estimates, on the other hand, are almost identical using the two methods. Tests of these hypotheses are contained in Panel B. The tests confirm that the order processing component from the Lee and Ready is statistically significantly different from that using the true classification for most

[7] For the analysis in Section 4.1, the transactions for which the true initiator could not be determined were simply eliminated. When the Lee study is replicated using only this classified subsample, the number of post-announcement periods with abnormal (small trade) buying decreases from three using the Lee and Ready method to one using the true classification (rather than from five to two with the full sample). It should be noted, however, that the elimination of this subset of the transactions could bias the results. For example, I find more abnormal buying in large trades in the subsample than in the full sample using both the Lee and Ready and the true classification.

[8] A number of alternative models of the spread have been proposed. Although I have not repeated the analysis using other specifications, I suspect that the results would be similar (particularly since the Glosten and Harris (1988) model is a special case of many of the other models).

Table 6

Panel A: Coefficient estimates

Method	Transitory component (c_0) in \$/share				Adverse selection component (z_1) in \$/1000 shares			
	Mean	25th %ile	Median	75th %ile	Mean	25th %ile	Median	75th %ile
Lee and Ready	0.0574	0.0519	0.0569	0.0622	0.0103	0.0010	0.0028	0.0080
True	0.0460	0.0387	0.0458	0.0530	0.0108	0.0009	0.0030	0.0079

Panel B: Hypothesis tests

Hypothesis	*p*-values			
	Mean	25th %ile	Median	75th %ile
H_0: Lee and Ready c_0 = True c_0	0.157	0.000	0.007	0.142
H_0: Lee and Ready z_1 = True z_1	0.624	0.463	0.709	0.846
H_0: Lee and Ready c_0 = True c_0 and Lee and Ready z_1 = True z_1	0.199	0.000	0.009	0.272

stocks. For example, the median *p*-value across the stocks in the sample is 0.007. The tests also confirm that the small differences in the adverse selection components are not significant. A joint test of the equality of both components is also rejected for most stocks (median *p*-value = 0.009). As expected, use of the Lee and Ready algorithm results in the overestimation of the costs of trading (or, equivalently, in the underestimation of price improvement).

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.

Barber, B., Odean, T., 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. Journal of Finance 55, 773–806.

Choi, H., Subrahmanyam, A., 1994. Using intraday data to test for effects of index futures on the underlying stock markets. Journal of Futures Markets 14, 293–322.

Ellis, K., R. Michaely, and M. O'Hara, 2000. The accuracy of trade classification rules: evidence from Nasdaq. Journal of Financial and Quantitative Analysis, forthcoming.

Glosten, L., Harris, L., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123–142.

Hasbrouck, J., 1988. Trades, quotes, inventories, and information. Journal of Financial Economics 22, 229–252.

Hasbrouck, J., 1992. Using the TORQ database. NYU working paper.

Lee, C., 1992. Earnings news and small traders. Journal of Accounting and Economics 15, 265–302.

Lee, C., Radhakrishna, B., 1996. Inferring investor behavior: evidence from TORQ data. University of Michigan working paper.

Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–746.

Petersen, M., Fialkowski, D., 1994. Posted versus effective spreads: good prices or bad quotes? Journal of Financial Economics 35, 269–292.

Ready, M., 1999. The specialist's discretion: stopped orders and price improvement. Review of Financial Studies 12, 1075–1112.

Schultz, P., Zaman, M., 1994. Aftermarket support and underpricing of initial public offerings. Journal of Financial Economics 35, 199–219.