A Liquidity Motivated Algorithm for Discerning Trade Direction

David Michayluk
University of Technology, Australia

Laurie Prather
Bond University, Australia

Most exchanges do not report trade direction thus researchers and traders must deduce whether a trade is buyer or seller initiated since this information is required to evaluate models of bid-ask spread components and to understand the market for immediacy. Algorithms that assign trade direction based on the proximity to bid or ask quotes are easily implemented but ignore information readily discernable from orders, changes in the quoted depth and subsequent price movements. Using the New York Stock Exchange Trades, Orders and Quotes database, systematic biases in existing trade direction algorithms are documented that can be rectified by recognizing that the impact on liquidity is the fundamental characteristic underlying order placement. Although this liquidity-based method is difficult to implement, it more closely captures the actual behavior of market participants (JEL : G10, G14).

Keywords: liquidity, trade direction algorithm, TORQ database, order placement

I. Introduction

As market microstructure systems come under increased scrutiny, the ability to accurately assess the strengths and weaknesses of alternative trading mechanisms has never been more important. The turmoil facing the New York Stock Exchange (NYSE) and its age-old specialist system is an example of traditional trading methods coming under examination. The NYSE controversy that may result in sweeping changes in the way securities are traded, illustrates the importance of addressing efficiency

* We thank Bruce Lehmann, Karyn Neuhauser and Henry Oppenheimer for helpful comments. All errors remain the responsibility of the authors.

(*Multinational Finance Journal*, 2008, vol. 12, no. 1/2, pp. 45–66)
Quarterly publication of the *Multinational Finance Society*, a nonprofit corporation.

DOI: 10.17578/12-1/2-3

issues regarding alternative trading structures.

Desirable qualities for the design of a market trading system are best price, speed and certainty of execution, and anonymity. Until recently, regulators have focused on best price and thus many academic studies have attempted to measure and compare the trading costs under different market microstructures.[1] Studies that estimate transaction costs and many market microstructure models rely on the classification of each trade as initiated by a buyer or a seller. In limit order markets without intervention where each transaction has a buyer and seller, simply comparing the transaction price to the inside-most buy and sell limit orders is sufficient to determine the trade originator. But in other, more complicated markets such as the New York Stock Exchange, specialists act as intermediaries and control the order flow. Orders are often altered before execution and do not have a transparent trade originator since the order may be changed to improve the price, to match orders, or to avert any adjustment in the quoted bid and ask price and depth. Because of these complicated trades, simply observing the price and timing of the trade is insufficient to identify the trade originator and existing trade direction algorithms acknowledge less than 100 percent accuracy.[2]

Examining the origin of each trade by tracing the trade to the original order introduces the concept of trade origination. Some orders provide liquidity to the market while others take away liquidity. This paper proposes using all the information surrounding the trade and order to determine the liquidity effect. By observing the liquidity motivation (*LM*) of each order, the price or time priority can be used with the order type and trades can then be classified based on their impact on liquidity. For example, orders that increase the availability of shares for

1. See for example, Christie and Schultz (1994), Huang and Stoll (1997), Bessembinder and Kaufman (1997).

2. Lee and Ready (1991) (*LR*) suggest the accuracy of their method is approximately 90%. Odder-White (2000) and Finucane (2000) suggest the accuracy is closer to 85%. Lee and Radhakrishna (2000) suggest the accuracy is 93% but this is after discarding approximately 40% of difficult to classify trades. Aitken and Frino (1996) and Theissen (2000) suggest that the accuracy is 75% on the Australian Stock Exchange and Frankfurt Stock Exchange, respectively. Ellis, O'Hara and Michaely (2000) (*EOM*) suggest that the accuracy is 80% on Nasdaq. Savickas and Wilson (2003) examine trade direction in the options market and find the quote rule to be 83% accurate, exceeding the *LR*, *EOM* and tick test accuracy levels of 80%, 77% and 59%, respectively, in this market. Boehmer, Grammig and Theissen (2007) suggest that inaccurate trade classifications bias estimates of the probability of informed trading and may severely affect the results of other empirical microstructure studies.

immediate execution, increase liquidity and do not necessarily result in a trade, while orders that reduce the availability of shares for immediate execution demand liquidity and result in a trade. This procedure examines the inputs to the trading process to classify resulting trades, and thus the resulting classifications are source-based rather than inference-based. Even though orders are used in this paper as one of the sources of information, changes in liquidity can be observed around trades to identify the trade direction, however this method would introduce some inferences.

In order to identify the severity of the bias, the trade direction classifications using the *LM* method are compared to those using the standard tick test rules, the Lee and Ready (*LR*) (1991) algorithm, and the Ellis, O'Hara, and Michaely (*EOM*) (2000) modification to the Lee and Ready algorithm. Estimates reveal that trades at the bid or ask provide the greatest consistency between the *LM* and other methods, but at other positions within the bid-ask spread trade directions are at best 70 percent consistent. All of the standard output-based trade direction algorithms result in effective spread estimates that are systematically biased when compared to the *LM* method. The *LR* technique and the *EOM* modification both overstate effective spreads while the tick test understates effective spreads. Thus, existing trade direction algorithms do not accurately measure transaction costs when trade liquidity is used to classify trade direction.

This paper is structured as follows: Section II examines the implied theoretical basis of trade-direction and how the liquidity motivation can be used to determine trade initiation classification. In Section III the consistency of the tick test, *LR* and *EOM* classification procedures are compared with the *LM* method. Section IV reports estimates of the effective bid-ask spread using each trade initiator classification technique. Finally, conclusions are detailed in section V.

II. Liquidity Motivation Trade Direction Classification Procedure

Before quote data were available, the tick test used only trade data to infer the direction of trading. This method uses the previous trade to gauge whether the current trade is at a higher price (and presumably a buy) or at a lower price (and presumably a sell). With the introduction of quote data, the proximity to either side of the bid-ask spread could be used to identify whether the trade was closer to the ask price (and presumably a buy) or closer to the bid price (and presumably a sell). The Lee and Ready (1991) algorithm uses the proximity of the transaction price to ask and bid quotes to determine trade direction, but if the trade is at the midpoint, then the tick test is used. For the Ellis, O'Hara and Michaely (2000) modification, trades at exactly the ask quote or bid quote are categorized as buys or sells, and all other trades are categorized using the tick test.

Classification of the direction of trading naturally developed from the availability of data. Since each order is either a buy or a sell, observing the trade would logically follow the order and a simple view of the order-trade sequence was employed by researchers. The complexity is apparent when examining the priority of the order. Market orders have time priority and they result in a trade quickly at whatever price, thus these are the trades typically observed at the extremes of the quoted bid-ask spread. On the other hand, limit orders have price priority and they do not necessarily result in a trade.

A natural improvement in determining trade direction is the use of order data. Order data can be linked to each trade and the resultant classifications determined from the source of the trade rather than an inference. Using the source of the trade requires an underlying definition of trade origination that may not be readily apparent in trade direction algorithms that use trade and quote data. This issue was identified by Odders-White (2000) where the timing of the order is used as the basis for determining the trade initiator. In this approach, whoever places the last order (buyer or seller) is assumed to be the trade initiator. While choosing a chronological definition because it is clear-cut and easy to apply, Odders-White recognizes that an alternative method would classify the trade initiator as the investor who demands liquidity. Lee (1992) and Peterson and Fialkowski (1994) identify the active side of the transaction as another theoretical basis to classify the trade initiator. While these approaches identify an underlying definition in determining trade direction, none provides a practical means to implement the theory.

A liquidity-motivation approach can use each order type to determine the liquidity influence. For example, if an order increases the number of shares available for immediate execution, this increase in the supply of shares increases liquidity and does not necessarily result in a transaction. Conversely, orders that remove shares for immediate execution reduce liquidity and result in transactions. A complication arises when any intervention in the market delays a transaction, even if only temporarily. Using liquidity as the basis for determining the trade direction ensures that any theoretical models that use the trade indicator are consistent with the liquidity function of the secondary market.

There are two liquidity-based definitions that can be used to assign trades. Either the liquidity motivation behind each order can be used or the actual liquidity effect of each trade can be employed. Using the actual liquidity effect is a weaker form that simplifies the classification process, but then the original intent of the order is lost and the trade direction classification does not capture the behavior of the market participants. Any alteration to the order by a market intermediary obscures the true intention of the order and the actions of the marketplace remain distorted.[3] While an admirable goal is the simplification of data and of the process of trade direction classification, not acknowledging market intervention is akin to throwing out information and leaving the black box of the trading process in place.

The second liquidity-based definition uses the liquidity motivation of the original order. Any intervention in the market will be better understood since this definition does not obscure the actions of the intermediary. For example, if an intermediary consistently alters a market buy order into a limit order (i.e., stopping or shopping the order), then this will be observable if the liquidity motivation definition is employed but not if the liquidity effect definition is used. The actions in the market are observable at each stage when the inputs to the trading process are the first source of data for determining the trade direction.

The mechanics of trade classification using orders is uncomplicated, since classification is dependent on liquidity demand. The procedures in assessing trade direction can be delineated by having each order separately assessed for its intended impact on liquidity. Altered trades such as market orders that are converted to limit orders will have an original order clearly indicating liquidity demand, so determining the intended trade direction is possible. It is precisely these types of trades that require order data and that may be misclassified using only trade and quote data. Werner (2003) examines orders that demand and supply liquidity and finds that liquidity-demanding orders often get price improvement. This process may make these trades harder to classify using conventional trade direction algorithms since market buy (sell) orders frequently execute at or below (at or above) the mid-quote. The

---

3. Order data itself is unnecessary if the liquidity effect is chosen as the underlying definition, since the liquidity effect is observable from trade and quote data.

relative proximity is often the main filter in trade direction algorithms, illustrating why more revealing data are important to accurately assess the trade direction.

Another complication arises when orders are matched and occur without the participation of the specialist. For example, when two non-marketable limit orders are crossed within the spread, then both orders could effectively provide liquidity but, because of their timing, the second one reduces liquidity.[4] In these situations, the second transaction is considered to be the trade initiator since it removes liquidity. Such transactions are among the difficult to classify degrees of immediacy to which Odders-White (2000) refers in choosing to examine only time priority. Lee and Radhakrishna (2000) avoid this issue by simply removing these trades from their examination. The ability to recognize these would-be non-marketable limit orders with order data allows for the determination of a new effective bid-ask spread. The first limit order acts to reduce the spread, and the second order becomes a marketable limit order that initiates a trade, thus reducing liquidity.

Crossed orders can also occur with market orders. If the specialist stops a market order and crosses it with another market order, then the resulting transaction is actually bi-directional since each order would have resulted in a transaction without the specialist's intervention. Thus, crossed orders can have one or two directions according to the liquidity motivation method. If only the liquidity effect were used, then crossed orders may be irrelevant since these trades may take place without the specialist altering the price or shares available in the market. Explicitly recognizing the one or two directions in crossed trades better reflects the economic actions of a quote driven market with specialist intervention.

Without any effect on the marketplace liquidity, crossed trades complicate the accurate measurement of effective spreads. A typical crossed trade of two market orders could have occurred as two trades – one at the bid and one at the ask price. In matching orders, the specialist may choose to improve both sides of the transaction such that no bid-ask spread is charged and the orders are crossed. If the crossing occurs at the midpoint then both traders are better off and the effective spread is zero. However, if the trade occurs at any price above or below the midpoint, then the average effective spread remains zero but is split

4. Non-marketable limit orders are limit orders with prices such that they cannot be executed immediately.

evenly with the price concession conceded by one trader awarded to the other trader. In studies of the effective bid-ask spread using NYSE data, this subtle point is often missed in effective spread calculations and the distance from each trade to the midpoint is measured. Only with order data linked to trade data can those unique transactions be identified in order to accurately assess the effective spread (and any biases introduced by not recognizing crossed orders). Crossed orders have no specialist participation as a buyer or seller and thus no captured spread for the specialist. There is also a net zero effective spread for the buyer and seller. Any study of the effective spread that does not control for the net zero effective spread of crossed orders will overstate the true effective spread. In this study, approximately 28 percent of the sample involves crossed trades.

An order can be split up or combined with other orders and resultant trades can be very complicated. For the liquidity motivation method, any combination of orders on one side of a transaction takes the highest degree of liquidity demanded, that is, if one order on that side is a market order, then the entire side is considered to behave as a market order. Likewise, if an original order is split up, each individual trade is considered separately in assessing trade direction. This method reflects the economic substance of the market that alters orders to be filled.

The systematic application of identifying orders that demand liquidity automates the algorithm. Unfortunately, order data are only available from automated systems such as Euronext Paris, where there are few complicated trades since there is no specialist involvement in order flow. As a guide for when order data are available on a more complicated exchange, the liquidity motivation method can be automated by determining whether each order would result in a trade when the order is placed. That is, if the order is a market order, or a marketable limit order, then it would be executed immediately and result in a trade since liquidity is demanded. Second, orders and trades can then be matched together and the sequence of orders could be observed to classify each trade. If a specialist intervenes and reduces the liquidity motivation of an order before the order is filled, the order still retains its liquidity motivation from the original order. If there are two liquidity supplying orders, the first liquidity supplier can be observed as to the impact on the bid-ask spread, and the second liquidity supplier will likely be liquidity demanding upon observing the first trade. In addition, some trades will be classified as bi-directional since orders from both sides would have resulted in a trade.

TABLE 1. Sample Summary Statistics

Name of Company Analyzed and TICKER	Number of Transactions in Sample	Total Number of Transactions in TORQ database	Sample Period	Liquidity Level
Alliance Capital Mgmt (AC)	347	1260	11/1 - 11/30	LOW
Acuson Corp (CAN)	403	2985	11/1 - 11/14	MEDIUM
Alcan Aluminum (AL)	928	4159	11/1 - 11/16	MEDIUM
Alexanders Inc. (ALX)	552	552	11/1 - 1/31	LOW
Boeing Inc. (BA)	375	18651	12/10	HIGH
Colgate Palmolive (CL)	427	6255	11/15-11/20	MEDIUM
Coachman Industries (COA)	300	300	11/1 - 1/31	LOW
Deltona Corp. (DLT)	190	190	11/1 - 1/31	LOW
General Electric (GE)	512	39384	11/7	HIGH
International Business Machines (IBM)	480	33856	1/7	HIGH

(*Continued*)

TABLE 1. (Continued)

Name of Company Analyzed and TICKER	Number of Transactions in Sample	Total Number of Transactions in TORQ database	Sample Period	Liquidity Level
Matsushita Electric Industrial Limited (MC)	281	281	11/1 - 1/31	LOW
Philip Morris Companies Inc. (MO)	528	37059	1/15	HIGH
Pittston Company (PCO)	599	3324	1/4-1/17	MEDIUM
Petroleum and Resources Corp (PEO)	915	915	11/1 - 1/31	LOW
Parker Hannifin Corp (PH)	494	3536	12/7 -12/20	MEDIUM
Potomac Electric Power Company (POM)	465	5655	12/10-12/14	MEDIUM
American Telephone and Telegraph (T)	975	39495	12/10	HIGH
Union Electric Company (UEP)	302	3037	1/10 - 1/18	MEDIUM
Wallace Computer Services Inc. (WCS)	530	1815	12/21 - 1/25	LOW
Wedgestone Financial (WDG)	256	256	11/1 - 1/31	LOW

Note: Stocks are divided into liquidity groups based on the number of shares traded in the entire TORQ database. Low liquidity stocks are those with less than 2500 trades, medium liquidity stocks have between 2500 and 10,000 trades and high liquidity stocks have over 10,000 trades.

In this paper, the liquidity motivation trade initiation classification is manually determined using the data files that comprise the New York Stock Exchange's Trades, Orders and Quotes (TORQ) database. Each of the transactions, quotes, orders, and audit files are combined into a continuous series and ordered for each security for each trading day. Care is taken to ensure the files are properly integrated.[5] This database encompasses a three-month period from November 1, 1990 to January 31, 1991 and includes the quotes, transactions, orders, and audit records for 144 stocks.[6]

The sample used in this paper consists of 20 stocks from the TORQ database chosen to include a range of liquidity levels. Our goal was to obtain approximately 500 trades for each stock that was examined to ensure observation of each possible combination of trades for each stock. If the stock traded frequently, a minimum of one day of trading was examined. If the stock traded infrequently then all trades were examined, or enough to approach the 500-transaction goal. A total of 9,859 transactions across 20 different stocks are classified over time periods ranging from one day to three months. Table 1 reports the stocks in the sample, the number of transactions classified for each company, the time period when these transactions take place, the total number of transactions in the TORQ sample during the three-month period, and a liquidity identifier based on the total number of transactions occurring in the sample period. Table 1 reveals that the sample includes stocks with volume ranging from 190 to 39,495 transactions during the three-month period. These stocks consist of five high, seven medium, and eight low liquidity stocks using the volume of trading to determine the hierarchy. A small sample bias is avoided by classifying an average of 493 trades for each company.

For six of the eight low liquidity stocks, all transactions are classified for the entire three-month period (November through January) of the TORQ database. The low liquidity stock transaction sample sizes range from a low of 190 transactions (*DLT*) to a high of 915 transactions (*PEO*). The seven stocks in the medium liquidity group have sample sizes varying from 302 transactions (*UEP*) to 928 transactions (*AL*) covering time periods between five and seventeen days. Finally, for the

5. Following Lee and Ready (1991), trades are compared with quotes that were in existence at least five seconds prior to the trade.

6. More recent order data from this market are not available. For a complete description of the TORQ database see Hasbrouck (1992).

five high liquidity stocks, the analysis uses only one day of transactions. The sample size for the high liquidity stocks ranges from a low of 375 transactions (*BA*) to a high of 975 transactions (*T*). For those stocks where classification did not include the entire three-month sample period, sections of data are extracted randomly from different time periods to avoid the possibility of a time specific event unduly influencing results.[7]

In order to classify each trade, the transaction and associated activity is examined to manually determine the direction of trade initiation based on whether the original order supplied or demanded liquidity as already outlined. As Lee and Radhakrishna (2000) recognize, some transactions do not have complete audit trails in the TORQ database. For those transactions that are not able to be conclusively classified, the order flow is observed and the most probable trade direction is determined. We followed the rules of determining liquidity motivation and also looked at the liquidity effect of the trade (subsequent changes in the quotes, price and depth). Unlike Lee and Radhakrishna (2000), who discard 34 percent of their sample due to a lack of order information, 15 percent of trades in this sample are identified with incomplete audit trails and results are presented with and without these trades. This procedure ensures that any conclusions drawn are not solely the result of trades classified without a complete audit trail. This time-consuming manual examination of all trades in the sample is necessary given these data but as order data become more readily available and linked to resulting trades, a liquidity motivation determination can be automated.

III. Trade-Initiation Algorithm Classification Accuracy

The *LM* method trade initiation classifications are compared with the classifications from three other classification techniques. The tick test method of trade initiation classification examines each trade relative to the previous trades. If the trade price increases (decreases) then the tick test classifies that trade as a buy (sell). Under this procedure, opening trades and subsequent trades that occur without a price change are not classifiable. In this sample, 9.2 percent of trades are not classifiable using the tick test.

---

7. For two of the high liquidity stocks (Boeing and AT&T), the same day was chosen in error. Results were constructed with and without each company and no bias was found in including both companies in the overall results.

TABLE 2. Comparison of the Liquidity Motivation Trade Initiation Classification With Time-Based Trade Initiation Algorithms' Classification

Position	Number of Trades	Percent of Total	TICK	*LR*	*EOM*
Above Ask	198	2.0	4.6%	51.0%	4.6%
Ask	3,337	33.8	82.2%	93.6%	93.6%
Between Ask and Midpoint	499	5.1	67.1%	67.1%	67.1%
Midpoint	2,313	23.5	60.1%	60.1%	60.1%
Between Midpoint and Bid	438	4.4	67.1%	69.6%	67.1%
Bid	3,058	31.0	77.9%	94.0%	94.0%
Below Bid	16	0.2	100.0%	100.0%	100.0%
Overall	9,859	100.0	72.7%	82.6%	81.6%

Note: The percentage of identical classifications is reported using the liquidity motivation method relative to the alternative methods. TICK refers to the tick test, *LR* refers to the Lee and Ready (1991) algorithm and *EOM* refers to the Ellis, O'Hara, and Michaely (2000) modification.
.

The *LR* technique requires trade information as well as the outstanding bid and ask quotes in order to classify trades as buyer or seller initiated. Transactions are classified based on their location relative to the outstanding bid and ask quotes. Using outstanding quotes in effect at least five seconds prior to a transaction, trades occurring at the bid (ask) are classified as sells (buys), while all other trades are classified based on their proximity to either the bid or the ask quote. Trades at the midpoint are classified using the tick test. The *LR* classification procedure results in no classification for transactions occurring at the midpoint of the outstanding bid-ask spread if the previous transactions were all at the same price. This lack of classification often occurs for transactions at the beginning of the day. In this sample, 1.9 percent of trades are not classifiable using the *LR* algorithm.

Finally, the *EOM* procedure requires both trade and quote information. Similar to the *LR* procedure, a trade at the ask (bid) price is classified as a buy (sell), while, all other trades are classified using the tick test regardless of their relative position within the bid-ask spread. No classification is possible if a trade occurs away from the quoted bid or ask price when there were no previous price changes. In this sample, 4.8 percent of trades are unclassifiable using the *EOM* procedure.

Similar to the other methods, when the *LM* method is used, the beginning of the day poses some difficulties since some audit data are missing. 16.9 percent of the trades with incomplete audit trails occur within the first half-hour of the trading day even though this period represents less than 8 percent of the trading day.

Table 2 compares the *LM* method classifications with the classifications from using the tick test, *LR* and *EOM* methods using the entire sample of 20 stocks. Between 73 percent and 83 percent of the *LM* method classifications are consistent with the classifications using the other methods. The tick test is 73 percent consistent while the *LR* and *EOM* are 83 percent and 82 percent consistent, respectively.

The classification consistency of the various methods is also considered based on the position of the trade relative to the bid, ask, and midpoint of the spread. For the full sample of 20 stocks, 65 percent of the trades occur at either the bid or the ask price. In general, for those trades, there is a fairly high level of consistency across the various classification techniques. For example, agreement with the *LM* method ranges from a low of 78 percent for the tick test classification of transactions at the bid to a high of 94 percent for *LR* and *EOM* classification of transactions at the bid. Additionally, while only 0.1

TABLE 3. Comparison of the Liquidity Motivation Trade Initiation Classification with Time-Based Trade Initiation Algorithms' Classification across Trades with and without an Audit Trail.

Position	Number of Trades	TICK	*LR*	*EOM*
Panel A. Trades with a Complete Audit Trail				
Above Ask	143	0.0%	46.2%	0.0%
Ask	2,980	82.8%	94.4%	94.4%
Between Ask and Midpoint	385	64.4%	66.8%	64.4%
Midpoint	1,756	59.4%	59.4%	59.4%
Between Midpoint and Bid	339	63.7%	71.1%	63.7%
Bid	2,741	78.6%	95.0%	95.0%
Below Bid	3	100.0%	100.0%	100.0%
Overall	8,348	73.5%	84.2%	83.0%
Panel B. Trades without a Complete Audit Trail				
Above Ask	55	16.4%	63.6%	16.4%
Ask	357	77.0%	86.8%	86.8%
Between Ask and Midpoint	114	76.3%	68.4%	76.3%
Midpoint	557	62.3%	62.3%	62.3%
Between Midpoint and Bid	99	78.8%	64.7%	78.8%
Bid	317	71.9%	84.9%	84.9%
Below Bid	13	100.0%	100.0%	100.0%
Overall	1,515	68.6%	73.8%	73.6%

Note: The percentage of identical classifications is reported using the liquidity motivation method relative to the other methods. TICK refers to the tick test, *LR* refers to the Lee and Ready (1991) algorithm and *EOM* refers to the Ellis, O'Hara, and Michaely (2000) modification.

percent of trades occur below the bid price, in these cases there is perfect classification agreement between all methods. Alternatively, when trades occur above the ask price there is far less agreement with the liquidity-based classification (only 4.6 percent for the tick test and *EOM* and 51 percent for *LR*). Approximately 23.5 percent of the trades occur at the midpoint and consistency between the liquidity-based method and the other methods is 60.1 percent for all methods since each method uses the tick test procedure for trades at the midpoint. Finally, in the 9.5 percent of the trades that occur between the ask price and the midpoint and between the midpoint and the bid price, the *LM* method leads to the same classification as all other methods for approximately two-thirds of the transactions. A general conclusion is that the existing trade initiator classification methods are most consistent with the *LM* method when trades occur at the bid and ask prices.

Odders-White finds that there is a systematic misclassification of transactions that occur at the midpoint, transactions that are relatively small, and transactions of stocks that are frequently traded or are highly capitalized. Since the *LM* method determines direction from the inputs to the trading process, there is no systematic susceptibility to misclassification of specific types of trades. The only potential erroneous classification is if order data are missing as in the case in 15 percent of this sample from the TORQ data. In future research using order data, proper matching should prevent any systematic misclassification.

Table 3 further examines the consistency between the classification methods by dividing the sample of stocks into two groups based on the completeness of the audit trail. Panels A and B report the results for trades with complete and incomplete audit trails, respectively. Approximately 85 percent of the trades in the sample have a complete audit trail. Confidence in the accuracy of the *LM* method classifications should be highest for these transactions and provide the best comparison base with the other methods.

Consistency between the *LM* method and other methods is slightly higher when a complete audit trail is available. This result is driven by the higher consistency of the large number of transactions occurring at either the bid or the ask in the complete audit trail subsample. For trades at and around the midpoint, the consistency between the *LM* method and other methods is slightly higher for trades with an incomplete audit trail. Importantly, these results reveal that the misclassification between the *LM* method and other methods is not confined to those trades without a complete audit trail.

Table 4 compares the classifications based on the number of originating orders. Single order trades represent 72 percent of the sample while 28 percent of the sample consists of trades with two originating orders. Those trades with only one originating order exhibit an overall consistency with the *LM* method of 71 percent, 82 percent, and 81 percent for the tick test, *LR*, and *EOM* methods, respectively. Trades at the midpoint reveal a much lower degree of consistency of 52.6 percent. Classification of trades between the bid or ask price and

TABLE 4. Comparison of the Liquidity Method Trade Initiation Classification with Time-Based Trade Initiation Algorithms' Classification across Trades with One or Two Originating Orders.

Position	Number of Trades	TICK	*LR*	*EOM*
Panel A. Subset of Trades with One Originating Order				
Above Ask	190	4.7%	51.6%	4.7%
Ask	2,523	83.2%	94.3%	94.3%
Between Ask and Midpoint	271	53.1%	55.4%	53.1%
Midpoint	1,427	52.6%	52.6%	52.6%
Between Midpoint and Bid	261	55.6%	59.8%	55.6%
Bid	2,381	77.2%	95.0%	95.0%
Below Bid	16	100.0%	100.0%	100.0%
Overall	7,069	70.7%	82.2%	80.7%
Panel B. Subset of Trades with Two Originating Orders				
Above Ask	8	0.0%	37.5%	0.0%
Ask	814	79.1%	91.7%	91.7%
Between Ask and Midpoint	228	83.8%	81.1%	83.8%
Midpoint	886	72.2%	72.2%	72.2%
Between Midpoint and Bid	177	84.2%	84.2%	84.2%
Bid	677	80.5%	90.3%	90.3%
Below Bid	0			
Overall	2,790	77.7%	83.7%	83.8%

Note: The percentage of identical classifications is reported using the liquidity motivation method relative to the alternative algorithm. TICK refers to the tick test, *LR* refers to the Lee and Ready (1991) algorithm and *EOM* refers to the Ellis, O'Hara, and Michaely (2000) modification.

the midpoint are only slightly more consistent. This disturbing observation indicates a systematic inconsistency between the *LM* method and conventional trade direction classification methods.

Overall, when orders that could have resulted in two separate trades are examined, the *LM* method is consistent with the tick test and *LR* classifications 78 percent of the time, and with the *EOM* test classifications 84 percent of the time. This subset is characterized by trades within the bid-ask spread that exhibit a higher degree of consistency between the *LM* and other methods than the subset of trades with one originating order.

In sum, for all trades within the bid-ask spread, there is a startling lack of consistency between the *LM* and other methods of trade initiation classification. Trades at the midpoint are only 60 percent in agreement; with other inside quote trades slightly more consistent but none above 70 percent. This finding reveals that the *LM* method and conventional methods are not similar and there may be biases in using conventional trade direction algorithms compared to using the *LM* method. One potential bias is examined in the next section.

IV. Effective Spread Estimates

Effective spread estimates using trade and quote data have recently generated some research interest. Peterson and Sirri (2003) use order data to calculate the bias in trading cost estimators resulting from inferences typical trade and quote users must make in lieu of order data. They find that trading costs are overstated by 17 percent and are systematically biased with the largest overstatement for small trades in large capitalization stocks. They caution that their analysis is limited to mostly retail orders sent to an auction market so there are many other types of orders that are not analyzed.

Werner (2003) examines all orders, including those excluded by Peterson and Sirri (2003), and is able to identify how different types of orders are systematically misclassified. The *LR* algorithm misclassified almost 30 percent of all market orders and some order types have even higher misclassifications. This finding suggests a drastic overstatement of spreads using the *LR* algorithm to classify trade direction.

Crossed trades pose a special problem when assessing effective spreads. For crossed trades on the NYSE, the specialist may choose to improve both sides of the transaction such that no bid-ask spread is charged. If the crossing occurs at the midpoint, then both traders are better off and the effective spread using conventional time-based trade algorithm methods is zero. If the crossing occurs within the bid-ask spread at any price above or below the midpoint, the combined effective spread remains zero, but the location of the trade within the spread affects the transaction costs of the two traders. Both traders' costs are better than if the quoted prices were paid, but compared to the midpoint, one trader's cost is negative and the other trader's cost is positive. In

TABLE 5. Effective Bid-Ask Spreads using the Liquidity Motivation Trade Initiation Classification and Time-Based Trade Initiation Algorithms' Classification

	Effective *BAS* (cents)	Effective Percentage *BAS* (%)
Liquidity	10.51	1.46
Motivation Method	(13.79)	(3.12)
TICK	9.18	1.14
	(7.69)	(2.19)
LR	15.25	1.96
	(16.94)	(4.21)
EOM	14.28	1.90
	(15.97)	(4.12)

Note: BAS is the bid-ask spread and the percentage bid-ask spread is measured as the bid-ask spread divided by the midpoint of the bid-ask spread. TICK refers to the tick test, *LR* refers to the Lee and Ready (1991) algorithm and *EOM* refers to the Ellis, O'Hara, and Michaely (2000) modification. Standard deviations are reported in brackets.

studies of the effective bid-ask spread using NYSE data, this subtle point is often ignored in effective-spread calculations, and the direction from the midpoint dictates the order direction and cost estimate. However, with order and audit data, these special transactions can be identified, the effective spread can be more accurately assessed and any biases caused by the trade initiation algorithm can be identified.

Table 5 reports the average spread measures for the entire sample using each trade direction algorithm. Columns two and three report results for the absolute effective spread and the percentage spread, respectively. The results suggest that spreads are lowest using the tick test with an effective spread of 9.18 cents and a percentage spread of 1.14 percent. The highest effective and percentage spreads of 15.25 cents and 1.96 percent, respectively, are found using the *LR* algorithm. Using the *EOM* procedure, the effective bid-ask spread is 14.28 cents and the effective percentage spread is 1.90 percent. While the tick test results in the lowest spread measures, the *LM* method provides measures that are lower than both the *LR* and *EOM* methods. The *LM* method results in an effective spread and a percentage spread measure of 10.51 cents and 1.46 percent, respectively.

The uniqueness of the *LM* method lies in its ability to identify crossed trades and this is only possible because of the richness of the TORQ dataset. Given that this approach is feasible using TORQ data, it is

TABLE 6. Effective Bid-Ask Spreads Across Liquidity Groups

	Effective *BAS* (cents)	Effective Percentage *BAS* (%)
Panel A. High Liquidity Stocks (2,870 Trades)		
Liquidity Motivation		
Method	6.01	0.13
	(2.36)	(0.11)
TICK	5.87	0.13
	(1.74)	(0.09)
LR	8.76	0.19
	(0.98)	(0.10)
EOM	8.76	0.19
	(0.98)	(0.10)
Panel B. Medium Liquidity Stocks (3,618 Trades)		
Liquidity Motivation		
Method	6.67	0.28
	(4.08)	(0.18)
TICK	7.98	0.34
	(4.61)	(0.21)
LR	11.50	0.48
	(4.88)	(0.23)
EOM	10.45	0.44
	(4.54)	(0.22)
Panel C. Low Liquidity Stocks (3,371 Trades)		
Liquidity Motivation		
Method	16.68	3.32
	(20.64)	(4.44)
TICK	12.31	2.48
	(10.96)	(3.08)
LR	22.59	4.36
	(25.54)	(6.09)
EOM	21.07	4.24
	(24.18)	(5.97)

Note: BAS is the bid-ask spread and the percentage bid-ask spread is measured as the bid-ask spread divided by the midpoint of the bid-ask spread. TICK refers to the tick test, *LR* refers to the Lee and Ready (1991) algorithm and *EOM* refers to the Ellis, O'Hara, and Michaely (2000) modification. Standard deviations are reported in brackets.

worthwhile to determine the inherent biases in the different trade initiation algorithms, since stocks more likely to benefit from specialist trade crossings will have a lower effective spread. This finding has implications for studies that compare the effective spread across different exchanges. For example, on the NYSE, according to the NYSE fact book, specialists were involved as a buyer or seller in only 30.2 percent of trades in 2001, meaning that almost 70 percent of trades were crossed.

In this sample, the medium and low liquidity stocks have 36.0 percent and 30.8 percent of their trades crossed with other traders, respectively. High liquidity stocks have 17.5 percent of their trades crossed. This difference between the stocks in different liquidity groups indicates that the effective spread overstatement may be more severe for medium and low liquidity stocks. Table 6 reports effective spread measures based on the three liquidity categories. Panels A, B, and C present spread measures for high, medium and low liquidity stocks, respectively. Qualitatively, the effective bid-ask spread results for the four different classification techniques are consistent with the overall results in table 5. In particular, both spread measures are lowest for the tick test, and the *LM* method, and highest for the *LR* and *EOM* methods, regardless of the level of liquidity.

The high liquidity stocks have effective percentage spread measures that are half the size of the medium liquidity stocks. The low liquidity stocks have shockingly high effective percentage bid-ask spreads that are about ten times larger than the medium liquidity stocks. In each of the categories, the *LM* effective spread measure is 20-30 percent smaller than the corresponding *LR* or *EOM* measure, indicating that the overstatement is not simply related to the number of crossed-trades but also to other classification differences.

V. Conclusion

Classifying trade initiation based on an order's liquidity motivation introduces a new, and potentially superior, method in the search for the proper trade initiation classification procedure. Existing methods have to infer direction from trades and quotes. This process is analogous to using the output from a black box to try to decipher the trading mechanism that converts orders to trades within the black box. A cleaner and more accurate procedure is to use the inputs to the process. But before order data are employed, it is important to clarify the theoretical basis for the trade-direction algorithms. Using an unambiguous definition, where the liquidity intent of the order is the primary criterion, enables a rigorous classification procedure. Additionally, ensuring that there is a theoretical foundation for assessing trade direction, the central assumption of the behavior of traders is consistent with the determination of the trade direction, many microstructure models, transaction cost estimations and the liquidity function of the secondary market.

Analysis of nearly 10,000 transactions reveals a number of striking patterns. The level of consistency between the *LM* method and other trade initiator classification procedures varies for trades within the bid-ask spread, with trades at the quotes having the highest degree of consistency. Within the bid-ask spread, about one third of trades have different classifications depending on the trade-initiation classification method. When this difference is combined with the recognition that crossed trades produce a zero effective spread, the difference in the calculated effective spread between the various trade initiator classification methods is substantial. The *LM* method provides consistently smaller measures than the *LR* and *EOM* method, irrespective of the liquidity level of each stock.

These findings concerning the effective bid-ask spread have far-reaching implications in comparisons of trading costs across stocks traded on the New York Stock Exchange and competing stock exchanges. The systematic overstatement of actual trading costs when crossed trades are ignored devalues the importance of the NYSE specialist in facilitating trading at a minimum cost. Approximately 28 percent of trades identified as crossed would have had an effective bid-ask spread measured under standard procedures of approximately 15 cents or 2 percent. The percentage of crossed trades and specialist intervention appears to have increased with less direct specialist participation in transactions since the TORQ data time period, thus, the overstatement of the true effective spread may be even larger today and most certainly varies substantially among stocks. The important finding from this research is that the inferences from using only trade and quote data have excluded the very information that identifies the comparative advantage of the NYSE specialist.

This research documents the importance of understanding order motivation as it relates to the effect on liquidity to accurately determine trade direction. Combining a clear theoretical basis for trade direction with sufficient data will ensure that empirical testing of market microstructure models is accurate and consistent with the behavior of traders. Although assumptions about time-preference in existing trade direction algorithms may allow the data to become more tractable and may facilitate empirical estimation of market microstructure models, any conclusions must be tempered by the obvious concern that the large number of improperly classified trades may affect estimated parameters.

References

Aitken, M. and Frino, A. 1996. The accuracy of the tick test: Evidence from the Australian Stock Exchange. *Journal of Banking and Finance* 20: 1715–1729.

Bessembinder, H. and Kaufman, H. 1997. A cross-exchange comparison of execution costs and information flow in NYSE-listed stocks. *Journal of Financial Economics* 46: 293–319.

Boehmer, E., Grammig, J. and Theissen, E. 2007. Estimating the probability of informed trading – Does trade misclassification matter? *Journal of Financial Markets* 10:26–47.

Christie, W. G. and Schultz, P. H. 1994. Why do market makers avoid odd-eighth quotes? *Journal of Finance* 49: 1813–1840.

Ellis, K., Michaely, R. and 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.

Lee, C. 1992. Earnings news and small traders: An intraday analysis. *Journal of Accounting and Economics* 15: 265–301.

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

Lee, C., and Radhakrishna, B. 2000. Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets* 3: 83–111.

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

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

Peterson, M. and Sirri, E. 2003. Evaluation of the biases in execution cost estimation using trade and quote data. *Journal of Financial Markets* 6: 259–280.

Theissen, E. 2001. A test of the accuracy of the Lee and Ready trade classification algorithm. *Journal of International Financial Markets, Institutions and Money* 11: 147–165.

Savickas, R. and Wilson, A. 2003. On inferring the direction of option trades. *Journal of Financial and Quantitative Analysis* 38: 881–902.

Werner, I. 2003. NYSE Order flow, spreads, and information. *Journal of Financial Markets* 6: 309–335.