The Accuracy of Trade Classification Rules: Evidence from Nasdaq

Katrina Ellis
Australian Graduate School of Management

Roni Michaely
Cornell University and Tel-Aviv University

And

Maureen O'Hara
Cornell University

April 2000

The authors would like to thank Dean Furbush, Tim McCormick, and Jennifer Drake of the NASD Economic Research Department for extensive help in providing the data set used in this paper. We thank the editor (Paul Malatesta), Jeffrey Harris, Soeren Hvidkjaer, Tim McCormick, and Paul Schultz for useful comments. We are particularly grateful to an anonymous referee for helpful suggestions. Please address correspondence to Roni Michaely (rm34@corenll.edu) or Maureen O'Hara (mo19@cornell.edu) at the Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853.

The Accuracy of Trade Classification Rules: Evidence from Nasdaq

Abstract

Researchers are increasingly using data from the Nasdaq market to examine pricing behavior, market design, and other microstructure phenomena. The validity of any study that classifies trades as buys or sells depends on the accuracy of the classification method. Using a proprietary data set from Nasdaq that identifies trade direction, we examine the validity of several trade classification algorithms. We find that the quote rule, the tick rule and the Lee-Ready (1991) rule correctly classify 76.4%, 77.66% and 81.05% of the trades, respectively. However, all classification rules have only a very limited success in classifying trades executed inside the quotes, introducing a bias in the accuracy of classifying large trades, trades during high volume periods, and ECN trades. We also find that extant algorithms do a mediocre job when used for calculating effective spreads. We propose a new and simple classification algorithm for Nasdaq trades that improves over extant algorithms.

The Accuracy of Trade Classification Rules: Evidence from Nasdaq

The availability of intraday trade and quote data has allowed research into a wide variety of issues in securities markets. A critical factor for many of these studies is the ability to determine trade direction. Who is buying and who is selling is an important element in determining the information content of trades, the order imbalance and inventory accumulation of liquidity providers, the price impact of large trades, the effective spread, and many other related questions. Unfortunately, commonly available high frequency databases do not provide information on trade direction. Consequently, empirical researchers have relied on trade direction algorithms to classify trades as buyer-or seller-motivated.

Three trade classification algorithms have been extensively used in microstructure studies: (1) the quote rule, (2) the tick rule, and (3) the Lee-Ready (1991) rule.[1] The quote rule classifies a transaction as a buy if the associated trade price is above the midpoint of the bid and the ask; it is classified as a sell if the trade price is below the midpoint quote. Trades executed at the midpoint are not classified. The tick rule classification is based on price movements relative to previous trades. If the transaction is above (below) the previous price, then it is a buy (sell). If there is no price change but the previous tick change was up (down), then the trade is classified as a buy (sell). The Lee-Ready procedure is essentially a combination of these two rules: first, classify a trade according to the quote rule (above or below the mid-point), and then classify the mid-point transaction using the tick rule. Given the reporting procedure on the NYSE, Lee and Ready also suggest comparing transaction prices to quotes reported at least five seconds before the transaction was reported.

How well do these algorithms work? This is difficult to determine without information on the source of the original order. What little evidence exists is virtually all based on NYSE data. Using the TORQ database, Lee and Radahakrishna (1996) report an overall 93% agreement between the actual order and the Lee-Ready algorithmic inference. Using the same data source, but a different selection criterion, Odders-White (1999) reports a success rate of 85% for the Lee-Ready algorithm. In a recent paper, Finucane (2000) uses the TORQ database to test the accuracy of several classification algorithms (tick, quote and Lee-Ready). He concludes that for

[1] For the derivation of these approaches see Hasbrouck (1988), Blume, Mackinlay and Terker (1989), and Lee and Ready (1991).

NYSE firms, the tick test and the Lee-Ready method have very similar performance accuracy in classifying trades. He also shows that the tick rule provides a better estimate of the effective spread than the Lee-Ready procedure. The one study done with non-NYSE data, Aitken and Frino's (1996) examination of the success of tick rules on Australian data, finds only a 75% success rate. There has been, to date, no study of the accuracy of trade algorithms using Nasdaq data. Given the increasing interest in, and analysis of, the Nasdaq market, the ability to interpret accurately Nasdaq data has obvious importance.

In this paper we investigate the applicability and accuracy of trade-direction algorithms to Nasdaq data. What facilitates our analysis is a unique data set containing Nasdaq trades identified with respect to trader identity and direction, as well as complete data on Nasdaq quotes. These data allow us to investigate the accuracy of standard classification techniques, as well to assess the performance of alternative approaches. Our analysis focuses on four questions: First, how successful are the classification rules for Nasdaq stocks? Second what type of transactions or market conditions are more likely to result in erroneous classifications? Third, what difficulties arise in applying trade algorithms to publicly available data such as the TAQ (trades and quotes) database? Fourth, is there a more accurate algorithm that would improve upon classifications for Nasdaq stocks?

Our findings are both reassuring and cautionary. We find that the standard classification rules for sorting trades into buys and sells are comparably accurate for Nasdaq trades as for NYSE trades. The quote rule successfully classifies 76.4% of the trades, the tick rule successfully classifies 77.66% of the trades and the combination of the two (labeled as the Lee Ready classification) is successful in 81.05% of the cases. (Odders-White reports success rates of 75%, 79%, and 85% respectively for NYSE stocks.) The success rate for classifying trades inside the quotes, however, is substantially lower, falling to approximately 60% for mid-point trades, and to only 55% for trades that are inside the quotes but not at the mid-point.

Examining further the misclassification of trades reveals a number of regularities. Large trades are less accurately classified as are trades in periods of rapid trading. These findings can be almost entirely attributed to the fact that large trades and trades during high volume periods are more frequently executed in between the quotes. We also find that trades executed on Electronic Communications Networks (ECNs) may be particularly prone to miscalculation for the very same reason: a much larger portion of those trades executes in between the quote. We

also find that over 4% of trades are executed outside the quotes, and that as much as 35% of the trades above (below) the ask (bid) are seller (buyer)-initiated trades. These trades pose severe problems for any trade classification algorithm. Finally, we show that, given Nasdaq reporting and trading procedures, there is a substantial delay in the reporting of Nasdaq trades on the standard TAQ database. This delay, however, does not substantially degrade the ability of extant algorithms to classify accurately Nasdaq trades as buys or sells.

Based on these findings, we develop a new trade classification algorithm that improves upon classification of trades away from the quotes, resulting in an improvement over existing algorithms when applied to Nasdaq data. For sorting trades into buys and sells, our algorithm achieves an overall success rate of 81.9%, a modest improvement compared to the Lee-Ready value of 81.0%.

However, the classification errors have important implications for the reliability of classification algorithms for some specific applications. In particular, we find that using the Lee-Ready algorithm significantly overstates the size of the effective spread in the Nasdaq market. Using the correct buy/sell indicator from the Nasdaq audit data, we calculate an average effective spread of 1.34%, whereas using the Lee-Ready classifications results in an effective spread of 1.96%. (Using NYSE data Finucane (2000) also reports a large over-estimate of the effective bid-ask spread when using the Lee-Ready method.) In calculating effective spreads, our algorithm shows a 10% improvement over the Lee-Ready approach. Given the greater accuracy and reliability of our classification algorithm, we believe it is a better approach for research involving Nasdaq trades.

The paper is organized as follows. In the next section, we describe the data used in our analysis. In Section 2, we test the accuracy of several alternative trade classification algorithms for Nasdaq trades. Section 3 then investigates when the algorithms fail, with particular attention given to the types of orders and to features of the trading process. In Section 4, we propose an alternative algorithm for analyses of Nasdaq data. In Section 5, we analyze the robustness of our proposed algorithm, and we explore several possible explanations for the seemingly anomalous behavior of trades executed outside the quotes. We then integrate our analysis of Nasdaq data with the TAQ database to examine reporting delays. In the paper's final section, we summarize our findings, we discuss their implications for trade classification, and we raise concerns

regarding the future ability to classify trades given the increasing propensity for trades to arise between the quotes and from ECNs.

1. Data

The sample contains 313 Nasdaq stocks traded between September 27, 1996 and September 29, 1997. Each of these stocks began trading following their initial public offering sometime during this time period. Consequently, for each stock the time series of data varies from between three months to twelve months. The transactions data were provided by the NASD, and they are taken from their Market Data Server (MDS).[2] Overall, the sample contains 2,433,019 trades and 627,370 quotes.

Because the sample contains newly traded firms, it is worth comparing its characteristics to those of Nasdaq stocks as a whole. Using information from the CRSP tapes, we compared our sample to the Nasdaq universe (4700 stocks) in terms of market capitalization and trading volume. While the sample firms have a lower market capitalization than the average Nasdaq stock during this period ($172m and $322m respectively), the median firm in the sample has a larger market value than the median Nasdaq stock ($112m and $66m respectively). This suggests that the sample has stocks of greater uniformity, and it does not have the extreme stocks in terms of size. Also, the turnover rate of the sample stocks is 0.83% per day, which is larger than is typical for a stock traded on Nasdaq (0.60%). This is not surprising given the very high trading volume immediately after the IPO. When we eliminate the first month of trading for each firm, the turnover mean (and median) of the sample is 0.60% (0.49%), which is not different than those for Nasdaq as a whole.[3] Overall, at least in terms of market value and turnover, our sample stocks seem to be quite representative of Nasdaq stocks.

Our proprietary data provide the bid quote, the ask quote, the transaction price, the trade volume, a trader identity code, and a buy/sell indicator.[4] The buy/sell indicator specifies the direction of the reporting party. If the seller reported the trade, then the trade would be called a sale regardless of the price at which it occurred. Thus, this indicator tells us the buyer and the

[2] Detailed description of NASD data is provided in Smith et al (1997).
[3] As reported in Section 5, we have repeated the entire experiment excluding the first month of trading for each stock. None of our results materially changes.
[4] The full MDS database has other fields, including what system was used for the transaction, e.g. whether it was a small order execution system (SOES) trade or via an electronic communications network (ECN). However, we did not get these fields with our data.

seller in the transaction, but not the trade direction *per se*. We can determine whether the trade was a buy or a sale by using the trader identity code in conjunction with the buy/sell indicator. The trader identity code is a four-letter code for every NASD member. When a non-NASD member trades (e.g. a customer trading with their NASD member broker), the trader identity code is blank. Thus, we can identify market makers and brokers by name, while the blank identity code merely tells us that a customer is trading.[5] There are three broad categories of trades that we use: transactions between market makers, between market makers and brokers, and between market makers or brokers and customers.

We are not able to directly sign trades between two market makers (16.6% of the trades), or two brokers (4.5%). For 3.5% of the sample, we know that the trade took place on an ECN because one of the trader identities attached to the trade is the ECN identity code. We cannot sign the ECN trades as we do not have the time the orders were placed with the ECN. Because ECNs function as limit order books, we would need to know the time each order was placed, and then use time priority to infer trade direction, with the later order being the trade motivator. (This is how Odders-White infers trade direction using the TORQ database.) Similarly, for the market maker – market maker and broker – broker trades, we do not know which party initiated the trade. In total, we exclude 24.6% of the sample that we can not accurately sign.

For the remaining 75.4% of the sample, knowing the identity of the trader and whether he was buying or selling allows us to determine directly whether the trade was buy-or sell-initiated. These are market maker – customer or broker-customer trades (26.2%), and market maker – broker trades (49.2%). For example, suppose that a trade is recorded as a sell by the market maker and the other party is a customer. This trade only occurs because the customer wants to trade: the market maker is the liquidity provider, and we categorize such a trade as a buy-initiated trade.[6]

---

[5] There are three levels of access to the NASD trading screen. First, there are market makers who enter and revise quotes, and execute and report trades. Second, order-entry brokers, who see, but cannot change, all the quotes. Order-entry brokers can enter orders to trade directly to a market maker or trade via electronic communications networks (ECNs). The third level are regional brokers who only see the best quotes and cannot directly enter orders. We have the identity of parties in the first two levels.

[6] Our time period straddles the new SEC mandated order handling rules, thus we need to consider orders clearing against limits directly placed by customers. The SEC rules were introduced during 1997. Stocks were phased into compliance via an implementation schedule of 22 waves that started on 20 January 1997 and ended October 13 1997. Initially, stocks were drawn from the top 1000 stocks ranked on median daily dollar volume. All other stocks were phased into compliance starting August 4 1997, with the majority being phased in during September 1997. The impact of including customer limit orders in the quotes is a reduction in the inside quote size (see Barclay, Christie, Harris and Kandel (1999)). Thus we may see a reduction in the number of trades occurring inside the spread as the

In the main part of our analysis, we use the market maker-customer and market maker-broker sample. This sample contains 1,833,729 trades that we use to directly test the success of the trading algorithms. We also repeated all of the tests using only the smaller sample of broker or market maker to customer trades (26.2% of all trades), and these results (reported in Section 5) are virtually identical.

Trade algorithms also typically use quote information. Our data provide the inside quotes (the highest bid and the lowest ask), and we have filtered this data to remove any obvious discrepancies (for example, crossed quotes where the ask is lower than or equal to the bid). After excluding 235 crossed quotes, we are left with 627,135 inside quote revisions.

Table 1 gives the distribution of all trades with respect to the prevailing quote. This table suggests that the market maker-broker sample trades (49.2% of the sample) is, in fact, not different from the market maker-customer (and broker-customer) sample trades (26.2% of the sample). These trades occur at all prices with very similar frequency as the customer trades: about 72.5 to 76% of trades in both samples execute at the quotes, 12 to 14.5% execute inside the quote, and approximately 5 to 6% execute outside the quote.

By contrast, we find that ECN trades (3.54% of the sample) have a different distribution: 34.62% of ECN trades execute inside the quotes, compared with 14.24% for the customer-based sample, while 30.51% execute at the midpoint, compared to 8.25% for the market maker-customer sample. Thus, 68.75% of ECN trades are not at the quote compared with 27.53% for the customer based sample.[8] Trades between two market makers appear to have some similarity to the ECN trades, with 40% of trades occurring at the midpoint or inside the spread.

These data suggest that our sampling approach of using the sub-sample of 76% of all trades is reasonable. However, we may be introducing noise because of trade occurring on an ECN between a market maker and a broker. We return to this issue further in Section 5.

2. Success of Trade Classification Rules

We now examine the accuracy of three trading rules: (1) the quote rule, (2) the tick rule, and (3) the Lee-Ready procedure (i.e. first classify by the quote rule and then by the tick rule).

---

order handling rules are introduced. In Section 5 we examine the impact of the introduction of order handling rules on the success of the trade classification algorithms.

[8] This may be due to the anonymity of trading via an ECN which results in a lower price impact and lower execution costs (see Huang and Stoll (1996)).

Using the quote rule, trades are classified as a buy trade if the transaction price is above the mid-point. Trades are classified as a sell if they occur below the mid-point. Trades executed at the mid-point can not be classified by the quote rule. Using the tick rule, trades are classified using the price movement prior to the trade. If the transaction price is above the previous price, then the trade is classified as a buy. If it is below the previous price, then it is a sell. If there is no price change, but the previous tick change was up (down), then the trade is a buy (sell). The difficulty of the tick rule as a stand-alone method is that it doesn't take into account the quoted prices. By only using the tick rule for midpoint transactions, we get the Lee-Ready algorithm. Thus, the Lee-Ready procedure (absent the five seconds rule, which we address later) simply combines the quote rule and the tick rule in the following way. First, classify all trades by the quote rule; then, classify the mid-point trades by the tick rule.

The success rates of the classification rules are given in Table 2. The quote rule correctly classifies 76.4% of all trades; the tick rule correctly classifies 77.66% of all trades; and the success rate for the Lee-Ready algorithm on our data is 81.05% of all trades. Of the three candidate algorithms, therefore, the Lee-Ready is the most successful, yielding an improvement of 3.4% over the tick rule alone and an improvement of 4.65% over the quote rule alone. These differences are statistically significant.

These results suggest that trade classification schemes on average provide reasonably accurate determinations for Nasdaq trades. The error rate of even the best scheme is sufficiently high, however, to introduce substantial noise into empirical analyses. This suggests determining more precisely when trade algorithms fail. In the next section, we address this issue by considering several potential factors: the proximity to the bid or ask, the trade size, time to previous quote update, and firm size. We focus our attention on the Lee-Ready algorithm since it is the most successful (and widely used) of the algorithms.

3.When Do the Trade Algorithms Fail?

3.1 Proximity to the bid and ask

The quote rule (and, by extension, the Lee-Ready algorithm) is very successful in correctly classifying trades executed at the bid and ask prices. As Table 2 indicates, 74.66% of all trades in our sample occurred at either the bid quote or the ask quote. The quote rule correctly

classifies 88.68% of these trades. Because the quotes represent dealers' offers to buy or sell at these prices, this high success rate should not be surprising.

A more interesting question is how well classification rules work for trades that are not at the quotes. Our sample has 20.4% of all trades occurring inside the spread, and 4.93% of trades occurring outside the quotes. By comparison, on the NYSE Lee and Ready (1991) find 31.2% of trades occurring inside the spread, and Odders-White (1998) finds 18.9% of such trades using the TORQ database. Trades occurring at the mid-point of the spread cannot be signed using the quote rule, but instead rely on the tick rule. Here we find that the algorithm correctly classifies 60.52% of all trades. Trades can also occur within the quotes but not at the mid-point. Using the Lee-Ready algorithm, the success rate of classifying these trades is a surprisingly low 55.23%. Finally, the success rate of classifying trades occurring outside the spread is only 64.77%. This is surprising since we would expect those trades to be more easily classifiable than even trades occurring at the quotes. (We elaborate on this issue further in section 5.4.)

In the reminder of this section, we examine several situations that may lead to this lower rate of success, but we note at this point that trades not executing at the prevailing quotes reveal serious deficiencies with extant algorithms.

3.2 Trade Size

Does the size of a trade affect the likelihood of correctly classifying it as a buy or sell? Table 3 shows the relationship between correct classification and trade size. Overall, there is a monotonic relationship: better classification for smaller trades. For example, trades of less than 200 shares are correctly classified 81.73% of the times compared with 77.85% for trades over 10,000 shares. This is in contrast to Odders-White's (1999) finding that accuracy is lower for smaller NYSE trades. However, a closer examination shows that the better classification of Nasdaq small trades is due to their being more likely to execute at either the bid or the ask quote. Only 51.7% of the largest trades execute at the quotes, compared to 80.31% of the smallest trades. Looking inside the quote, we find 9.3% of the small trades (less than 200 shares) executing there, as compared with 23.85% of the largest trades (more than 10,000 shares). Likewise, while 10.16% of the large trades occur outside the quote, only 4.94% of the small trades are executed outside the quote. Conditioning on the location of the trade, we find trade classification accuracy increases for larger trades, however, the overall accuracy is lower.

3.3 Slow vs. Rapid Trading Times

Table 4 shows that classification success falls during rapid trading (Panel A), and following a quote change (Panel B). When trades occur less than 5 seconds apart, 72.58% of the trades are correctly classified, compared to 88.77% of the trades when they occur more than 5 minutes apart.[9] Interestingly, there is not much variation in the probability of trades executing inside the quote as a function of the intensity of the trade. This is not the case for trades executing outside the quotes. We find that 9.26% of trades happen outside the quote when in periods of rapid trading, as compared to 1.3% in period of slow trading. The opposite picture emerges for trades at the quotes. A trade is more likely to execute at the quote during slow trading periods (81.76%) relative to rapid trading period (69.56%). This behavior may again reflect features specific to a dealer market. When trading is rapid, dealers may not update quotes as quickly because orders are arriving. This can result in a dealer trading at stale prices relative to his updated quote (this phenomena can also result in a crossed market in which the best bid is above the ask price).

Our findings here suggest that the higher success rate in classifying trades during slow trading periods is not due to the speed of trading, but rather to the higher incidence of trades at the quotes. Nonetheless, our findings do suggest caution in interpreting trade data in times of high volume. Likewise, as indicated in Panel B, trade classification is more successful the longer the time delay since a quote update. Again, this is due to more trades executing at the quotes the longer the time since a quote change. However, trades within 5 to 90 seconds of a quote change have a lower classification success than trades immediately after a quote, and this is due to lower success for trades at the quote.

3.4 Multivariate Analysis

Given the influence on trade classification accuracy of the interaction of the proximity of the trade to the quote with trade size, trading speed and quoting speed, a multivariate analysis may highlight the residual importance of such factors above and beyond the proximity of the trade to the quote. To address this issue, we ran a logistic regression where the binary dependent variable, y (1 if successful classification, 0 otherwise) was transformed via the logistic transformation,

---

[9] Odders-White has noted the higher rate of errors for trade algorithms during heavy trading for NYSE data.

$$\log\left(\frac{\Pr(y=1)}{\Pr(y=0)}\right)=\overline{\beta}'\overline{x}+\varepsilon\,. \tag{1}$$

As independent variables we included dummy variables representing the location of the trade relative to the quote (at, outside, inside and mid-point) as well as the trade size (in thousands of shares). Other independent variables we include are the firm size (the average market capitalization over the sample period in tens of millions of dollars);[10] the time since the previous trade in minutes; and the time since the previous quote in minutes.[11] The logistic regression gives the relative importance of each variable in classifying trades correctly via the modified Lee-Ready rule. Because of the logit transformation, the regression coefficients cannot be interpreted in the standard way as the marginal effect of the independent variable on the probability of successful classification. Instead, we can calculate this effect via

$$\frac{\partial \Pr(y=1)}{\partial x}=\frac{\partial\Lambda(\hat{\overline{\beta}}'\overline{x})}{\partial\hat{\overline{\beta}}'\overline{x}}\frac{\partial\hat{\overline{\beta}}'\overline{x}}{\partial x}=\Lambda(\hat{\overline{\beta}}'\overline{x})\left(1-\Lambda(\hat{\overline{\beta}}'\overline{x})\right)\hat{\beta} \tag{2}$$

where $\Lambda(\hat{\overline{\beta}}'\overline{x})=\dfrac{\exp(\hat{\overline{\beta}}'\overline{x})}{1+\exp(\hat{\overline{\beta}}'\overline{x})}$ is the cumulative logistic distribution function and $\hat{\overline{\beta}}$ is the vector of parameter estimates and $\overline{x}$ is the vector of the means of the independent variables.

The results are reported in Table 5. The regression shows trade size, firm size, trading speed, and quoting speed are each less significant in determining the probability of correct classification than is proximity to the quotes. The probability of correct classification increases with trade size, decreases with firm size, increases with the time between trades (less rapid trading), and increases with the time between a quote update and a trade. However, the relationship is slight: the slope value shows that the marginal effect of these variables is minimal. Proximity to the quotes is the most important factor in affecting classification accuracy. The slope values suggest that trades occurring at the midpoint, the quotes or outside the quotes result in a 4%, 27%, and 7% respective increase in the probability of being correctly classified over trades inside the quotes. Finally, to check for stability, we also estimated the parameters of the

[10] Univariate analysis of market capitalization with trade classification accuracy showed better accuracy for small firms, due to more trades executing at the quote. We also analyzed the effect of the time of day, and there was no variation in trade classification accuracy.

[11] We performed univariate analyses with these variables, and found the proximity of the trade to the quote dominated the effect of firm size or trading and quoting rapidity.

logistic regression using the first half of the data. Our results suggest that the parameters are generally reliable, and that the characteristics that determine correct classification are the same in different samples.[12]

4. An Alternative Trade Classification Algorithm

Our analysis has demonstrated a number of empirical regularities with respect to trade classification of Nasdaq trades. While we find that standard algorithms applied to Nasdaq trades compare quite favorably to NYSE-based results, approximately 20% of all trades on Nasdaq are incorrectly classified. This is due to a number of factors, with the most important being the difficulty of assigning direction for trades within the quotes. We have also demonstrated a problem unique to Nasdaq, in that trades occurring outside of the quotes perform poorly with respect to trade classification.

What, then, is the best trade classification approach for researchers to take when working with Nasdaq trades? Although extant algorithms are adequate to the basic job of sorting trades, our work suggests that a refinement to the extant methods of classifying trades will do even better. Specifically, the difficulties of assigning non-quote trades suggest relying more on tick rules than on quote rules. Hence, we propose the following alternative trading algorithm:

> ***All trades executed at the ask quote are categorized as buys. All trades executed at the bid quote are categorized as sells. All other trades are categorized by the tick rule***.

Table 6 reports the success rate of this alternative classification rule and compares it to the Lee-Ready algorithm.[13] The classification accuracy for trades executed inside the quotes increases from 55% to 61%, a significant improvement. There is also an improvement, but to a lesser extent, for trades outside the quote: from a 64.8% success rate under the Lee-Ready procedure to 65.8% under the alternative procedure. This improvement translates to an overall

[12] With a cutoff value for the estimated probability of 0.5, 80.4% of the trades are correctly classified. The out-of-sample predictive ability of the regression is tested on the remainder of the data by fitting the regression coefficients. Using a cutoff of 0.5, (i.e. if the estimated probability is above 0.5 then the trade is correctly classified), the error rate is 26.9%, dictating that 73.1% of trades are correctly classified.

[13] Given the logistic regression results, we do not find any reason to adjust the algorithm for time of day issues, nor do we feel it advisable to treat order size differently.

success rate of 81.87% for our candidate algorithm, compared to 81.05% for the Lee-Ready algorithm.

While this may seem a relatively modest improvement, a more compelling case for the proposed algorithm arises in specific applications. In particular, one important application of trade classification algorithms is in the computation of effective bid-ask spreads. Effective bid-ask spreads measure the difference between actual trade prices and quotes (see Petersen and Fialkowski (1994)). When trades routinely transact between the quotes, this measure will give a smaller, and more accurate, estimate of transactions costs than do posted bid ask spreads. The effective spread is calculated as

Effective spread = 2I[transaction price – midpoint price]

Where I is an indicator variable that equals one for buy trades and negative one for sell trades, and the midpoint is the average of the bid and ask prices. Obviously, correct classification of buy and sell trades is crucial in calculations of the effective spread.[14]

For every market maker – customer, broker – customer and market maker-broker trade in the sample we calculate the effective spread using the previous quote. Using the correct buy/sell classification from the NASD audit trail, we calculate the average effective spread as $0.1997, or 1.34% of the price, as compared to the average posted bid-ask spread of $0.3363, or 2.33% of the price. Such dramatic differences between effective and posted spreads testify to the complexity of accurately measuring transactions costs in dealer markets.

Since audit trail data is not typically available to researchers, we now calculate the effective spreads using the Lee-Ready algorithm. We find an average effective spread of $0.2909, or 1.96% of the price. The surprisingly high error in this classification mirrors that found by Finucane (2000) in his analysis of NYSE trades, suggesting that the Lee-Ready algorithm is inaccurate in measuring effective spreads in both Nasdaq and the NYSE. This is largely due to the poor performance of the Lee-Ready algorithm in classifying trades between the quotes. Our suggested algorithm does much better. We find an effective spread of $0.2681, or 1.80% of the price. While still an overstatement of the actual cost, our measure is a 10% improvement over the estimate found using the Lee-Ready measure, a direct result of using the

[14] Some researchers use a simpler definition of effective spread: twice the absolute difference between the transaction price and the midpoint. Thus, all trades above the midpoint are implicitly assumed to be buys, and all trades below the midpoint are assumed to be sells. When the Lee-Ready classification is used, the two effective spread measures are identical.

tick rule for non-quote trades rather than assuming that every trade above (below) the midpoint is a buy (sell).

Our proposed algorithm thus provides a more accurate method for classifying Nasdaq trades. A natural concern is whether this approach is robust to factors such as different sampling techniques. A second issue is how reporting delays influence classification accuracy. In the next section, we address these concerns and the larger question of how well our proposed algorithm works when applied to publicly available databases.

5. Robustness

5.1 Trading beyond the IPO month

Our sample consists of stocks entering the market through IPOs. A natural concern is that trading surrounding this event may differ from "normal", thereby biasing our classification results. To test whether trade classification algorithms work differently in initial trading than in "normal" trading, we repeat the analysis excluding the first month of trading. Restricting the analysis to market maker – customer, broker – customer or market maker-broker trades after the IPO month yields a sample of 1,380,271 trades.

Table 7 shows that the distribution of trades is not different from the sample including the IPO: 4.9% of the sample is outside the quotes, 74.3% at the quotes, 13.4% within the quotes but not at the midpoint and 7.4% at the midpoint. Overall the quote rule, tick rule and Lee-Ready rule correctly classify 75.3%, 76.6% and 79.8% of the trades. Our new algorithm significantly improves upon these results by correctly classifying 80.8% of the trades. The rate of success is again much higher at the quotes (87.41%) than away from the quotes, with a 60.56% success rate at the midpoint, and a 61.55% success rate within the spread but not at the midpoint. We repeated the univariate analyses comparing trade classification success with trade size, time between trades and quotes and firm size, and we found identical results. The logistic regression reported in Panel B of Table 7, supports the better classification at the quotes, midpoint and outside the quotes. Overall, our conclusions support the hypothesis that trade classification rules work with similar success in normal trading periods.

5.2 Market Maker – Customer Trades

A possible problem in our data is that some trades between market makers and brokers may have actually occurred on an ECN, but were not reported as such. This matters because ECN trading essentially involves limit orders, and the direction of the active party may be difficult to infer. In particular, in signing a market maker-broker trade, it could be that the broker's order was placed on the ECN first, acting as the liquidity provider, and the market maker hitting the broker's order initiated the trade. We have excluded ECN trades from our sample, but any erroneous reporting will introduce noise into our analysis.[15] In general, we would expect in a dealer market that market makers are the liquidity providers and brokers the liquidity demanders, and that is the intuition underlying our simplifying assumption. Because the market maker – customer trades do not have this potential problem, a simple check on our analysis is to redo our accuracy results using this restricted sample.

This new sample includes 636,637 trades, or approximately 26% of the initial sample. Overall, the success of the trade classification algorithms improves slightly to 77.8%, 80.2% and 83.0% for the quote, tick and Lee-Ready rules. Our new tick rule first-quote rule second algorithm again successfully classifies trades at the rate of 84.2%. The success rate is highest for trades at the quotes (91.2%) compared to at the midpoint 64.8% (Panel A, Table 7). A logistic regression (reported in panel B) confirms this pattern: trades at the quotes, midpoint, outside the quotes are more likely to be correct than trades within the spread. These results suggest little difference in classification accuracy compared with the larger sample including the market maker – broker trades.

5.3 Reporting delays for trades

Reporting conventions may significantly affect the success rates of trade classification rules. On the NYSE, for example, the median delay in reporting trades is five-seconds. The incorporation of this five-second delay is one of the distinctive features of the Lee and Ready (1991) algorithm. How reporting delays affect Nasdaq trades is not clear, but it can be established using our data set. Our data are drawn from the NASD audit trail, and so the transaction time given is generally accurate. For users of publicly available Nasdaq trade-by trade data sets (such as TAQ), however, reporting delays may arise. These could arise, for

[15] Including ECN trades should add noise but not bias if the broker is equally likely to make or take trades on an ECN. Over our time period, ECN trades are only a very small part of the market for these stocks (approximately 4%), and so any effect should be small in any case.

example, because of the different execution systems available to brokers and market makers, and because of delays between execution time and reporting time. Thus, an important question for users of publicly available databases is how to adjust for any reporting delays.

The transaction times on the NASD audit trail are not perfect. A trade always appears with the time that the trade was entered into the Automatic Confirmation Transaction Service (ACT). In addition, if a trade occurred via an automated trading system such as SOES, SelectNet, ACES (Advanced Computerized Execution System), then the actual execution time is also reported and appears on the NASD audit trail. Also, if a trade is reported later than the allowed 90-second delay, then the actual execution time must be reported, and it also appears on the NASD audit trail. The result is that transaction times on the NASD audit trail are a mix of ACT times (if no alternative execution time is reported) and the more-accurate execution times (for automated trades or late-reported trades). By contrast, times reported on TAQ are based solely on the ACT times.

To measure the difference between the transaction time and the time reported on the NYSE TAQ database (the most widely used Nasdaq trade-by-trade database) we match trades in on our database with trades on TAQ. For the 313 stocks between 27 September 1996 and 29 September 1997, there were 2,479,093 trades on the NASD database, and 2,494,926 trades on TAQ. Trades were matched on size and price. If multiple trades match on both size and price, then time priority is used. For example, if there are two trades for 100 shares at $12.50 on TAQ, but only one on NASD, then the one closest in time to the NASD trade is matched. The matching rate is 96.85% for trades and 98.27% for quotes.

Table 8 gives the distribution of reporting delays for trades (panel A), and for quotes (panel B). The data clearly show evidence of two trade types: those with a one second reporting delay (12.4% of trades), and those with a reporting delay of 15 or 16 seconds (58.6% of trades). In total, 93.2% have a reporting delay of up to 16 seconds: the mean delay is 11 seconds and the median reporting delay is 15 seconds. This is much shorter than Blume and Goldstein's (1997) finding that the median reporting delay for transactions of NYSE stocks on Nasdaq is 31 seconds. By contrast, there does not appear to be a significant time delay in the reporting of quotes. We find that 88% of all quotes are reported within plus or minus one second of the time on the NASD audit trail, with a median report delay of zero seconds.

The reporting delays for trades and quotes suggest that inferences from the TAQ database may be prone to errors due to trades matched with the wrong quote. To measure the extent of this problem, we calculated the number of quote updates between the actual trade time (NASD data) and the reported trade time (TAQ). This is reported in columns 3-5 of Panel A in Table 8. Overall, only 10.93% of trades have a quote update in this interval, limiting the problem of referring to the wrong quote to, at worst, only a fairly small fraction of all trades.

We can calculate how much the trade reporting delay affects the Lee-Ready and tick-quote trade classification algorithm by matching trades on the TAQ with our market maker-customer sample trades. Of these trades, we matched 1,751,792 trades of 1,833,729 trade, or 95.5%. Table 9 reports the success of the Lee-Ready algorithm and tick-quote rule for these trades. Overall the classification success increases from 82.74%[18] to 83.74% when the TAQ trade times are used instead of the actual trade times. If trades are advanced by 15 seconds (adjusting for the median reporting delay), then we correctly classify 82.73% of the sample. We conclude that incorporating reporting delays into the classification algorithm does not improve their accuracy for Nasdaq data. Consequently, we recommend using the TAQ time with no delay when classifying Nasdaq trades from TAQ data.

5.4 Classifications of trades executed outside the quotes

A feature peculiar to dealer markets, in general, and to Nasdaq, in particular, is the large fraction of trades occurring outside of the posted quotes. We find 4.93% of the trades in the sample execute outside the spread, a figure nine times greater than the 0.5% of NYSE trades reported in Lee and Ready (1991).

Surprisingly, the classification success for trades outside the quotes is significantly lower than it is for trades inside the quotes. One would expect that trades above the ask (below the bid) would almost always be initiated by a buyer (seller), reflecting, for example, the premium that a large buyer is willing to pay for liquidity. We find, however, that the success rate of classifying trades occurring outside the spread is only 65.8%, implying that a large portion of these trades are actually buyers buying *below* the bid and sellers *selling* above the ask. This result mirrors the

[18] This overall success rate differs from the previous rate (81.8%) as we are only the using the subsample of matched trades.

findings of Bernhardt, Dvoracek, Hughson, and Werner (1999) who find that on the London Stock Exchange large trades actually receive better execution than do small trades. Why this occurs on the Nasdaq is not obvious, but four hypotheses seem most relevant.

One is simply that this divergence is specific to our data sample: 11.52% of the trades outside the quotes occurred on the IPO day, which may be a unique event. A second possible explanation is that reporting rules may introduce technical problems in interpreting data. The third is that the Nasdaq market structure of multiple dealers creates situations of multiple effective bids and asks at times of rapid and large volume. The fourth potential explanation is that features specific to a dealer market may result in trades outside the spread representing significant price improvement for the trader.

The unique feature of out data sample is that it includes trading around the IPO. As noted before, a significant portion of outside trades occurs on the IPO date. In fact, 6.5% of all trades on the offer day fall into this category. But what is so unique about the first day of trading? Two features seem most important. First, there is an extremely high volume of trade (Ellis, Michaely, and O'Hara (2000) report an average 60% first day turnover for the stocks in this sample). Second, the trading frequency is intense; over 50% of all trades on the first day occur within 5 seconds of the previous trade. Difficulties with order executions on Nasdaq on IPO days are not uncommon, with some traders buying above the inside quotes and others selling below. These execution difficulties largely reflect the logistical difficulties of large volumes being handled by multiple dealers in newly issued securities.[19]

However, even excluding the first month of trading (see Table 7), 4.89% of the trades occur outside the quote and the classification success of these trades is not any better: only 64.48% are correctly classified. Overall, this suggests that the IPO day is not the reason for either the large number of trades outside the quotes, or the lack of success in classifying outside-the-quote trades.

We now consider the issue of technical or reporting problems. Porter and Weaver [1998] present evidence that late reporting of trades is a more common occurrence on the Nasdaq than it is on the NYSE. We do not have a condition code for late trades in our data set, but we can

[19] Such difficulties are the subject of "Nasdaq's Problems with Trading Snags May Make it Vulnerable to Competition", Wall Street Journal, June 10, 1999. The article focuses on a recent IPO (theGlobe.com) in which some investors bought at the opening from Mayer & Schweitzer at $96, while other traders were simultaneously

investigate this issue by matching our trades with trades in the TAQ data, which does indicate a flag for late reports. We find 13.7% of the trades outside the quotes match a late-reported trade on TAQ.[20] Thus, late reporting is a significant factor in explaining the prevalence of trades outside the quotes.

But even the comparison to the late reported trades may not indicate the full extent of the problem. As noted before, trades may be reported up to 90 seconds after the transaction time, without being regarded as a late-reported trade. If there is a quote change between the execution time and the reporting time, then we may be overstating the fraction of trades that occur outside the spread. For the whole sample, 29.93% of all trades occur within 90 seconds of the previous quote, and so these could be labeled ambiguous trades. However, for trades outside the quote, 62.26% are ambiguous. So it is possible that these trades are reported with a delay and if quotes have moved during this time period, then what is recorded as an outside-the-quote trade is simply misclassified.[21]

To examine the extent of this potential problem, we compare the trade price to all the quotes in the 90 seconds prior to the trade, and examine whether the trade price was at or within any of the prior quotes. We find that 67% of the ambiguous trades occurred outside the spread for all of the quotes within a 90-second time frame. This suggests that for 33% of the ambiguous outside-the-quote trades, the current quote is incorrect. However, most ambiguous outside-the-quote trades simply occur outside the spread.

Although these technical issues do explain a fraction of the outside-the-quote trades, they are not the full explanation. If we exclude all the late-reported trades, and exclude the 33% of ambiguous trades that would be reclassified if we used a different quote rather than the current quote, we are still left with 3.38% (instead of 4.93%) of the whole sample being outside-the-quote trades. This is still six times the frequency reported on the NYSE in Lee and Ready (1991). Furthermore, only 67.28% of these are correctly classified with our algorithm compared to 65.82% for the whole sample. Thus, although correcting for technical issues may reduce the

---

selling to Bear Stearns at $90. The article notes that "the controversial opening illustrates the structural problems plaguing Nasdaq".

[20] Just under a quarter of these trades (24%) occurred on the IPO day, a finding consistent with our earlier discussion of the relatively chaotic trading conditions often found on the offer day.

[21] However, 90 seconds can be a considerable delay, and we find that 62% of the ambiguous trades have at least two quotes within the preceding 90 seconds, and 25% have at least 5 quote updates during this time. Thus, we cannot easily identify the correct quote.

frequency of outside-the-quote trades, the lack of success in classifying these trades remains a puzzle.

The third possible explanation for the outside-the-quote trades, and mediocre classification of these trades, is that in times of rapid trading, multiple effective bids and asks may prevail in a multiple dealer market. Indeed, the evidence in Table 4 indicates that whenever trading is rapid, the frequency of outside-the-spread trades is much higher and the classification success is lower. For example, when trades occur less than 5 seconds apart, 9.26% of the trades are outside the spread and only 62.97% of these are correctly classified (last column of Table 4). Conversely, when trades occur more than 300 seconds apart, outside-the-spread trades occur only 1.3% of the time, and 73.24% of these are classified correctly. In a multiple market maker system like Nasdaq, it is possible that, in times of high volume, information is fragmented across dealers, and multiple effective bids and asks can prevail. Unlike a NYSE specialist, a Nasdaq market maker cannot see all the order flow in periods of rapid trading, but rather sees only his own orders. This can result in market makers relying on (and trading at) their quote rather than the inside quote during fast trading, and thus many more trades occurring outside the spread.

The fourth factor accounting for outside-the-quote trades may be related to inventory consideration by market markers. One explanation that immediately comes to mind is trade size. However, only 20.1% of the trades outside the quotes are for between 1,000 and 10,000 shares, and 6.4% are for amounts greater than 10,000 shares. Thus, the vast majority (74%) of trades outside the quotes are actually small trades. The success rate of the Lee-Ready algorithm is highest for trades over 10,000 shares (71.74%), suggesting that customers are paying for liquidity when they transact large amounts. For trades between 1,000 and 10,000 shares, however, the trade classification success drops to between 60 and 67%, suggesting that market makers are giving these customers better prices than the prevailing quotes. This price behavior may reflect features specific to a dealer market. In particular, Bernhardt, Dvoracek, Hughson, and Werner (1999) suggest that in dealer markets relationships between traders and dealers can result in dealers providing their regular customers with "better deals". When such enduring relationships are important, these authors argue that trades will tend to be larger, resulting in larger traders garnering better price execution. Such behavior is not observed on specialist markets or in open limit order markets which by definition lack the relationship feature common to dealer markets.

On the other hand, for small trades outside the spread, the trade classification success of 64 to 67% suggests that although 33 to 36% are receiving executing at prices better than the inside quotes, most are executing at worse prices.[22] This seems at odds with best execution requirements, but cannot be explained away by data errors such as late-reported trades, or comparing to other quotes that fall within 90 seconds prior to the trade. However, over 55% of these small outside-the-spread trades occur within 5 seconds of the previous trade (and 80% occur within 30 seconds of the previous trade), which again suggests that in times of rapid trading, the multiple market maker system may create situations of multiple effective bids and asks.

Whatever the explanation, from the perspective of trade classification, these results suggest that trades outside the quotes are likely to degrade the accuracy of any trade classification algorithm. In particular, for 36.2% of trades outside the quotes, trades are sells above the ask and buys below the bid, the opposite of what would be expected (or predicted by a trade direction algorithm). Our results suggest that trade classification during periods of active trading, and particularly on IPO days, will be prone to greater error.

5.5 Impact of the Order-Handling Rules

Our sample period includes the introduction of the new SEC order handling rules. Did the changes in trade handling rules affect the distribution of trade execution relative to the quotes, and did they affect the accuracy of classification rules? Although we do not know the specific date of compliance for the individual stocks in our sample, we can separate the stocks into a pre-compliance period and a post-compliance period. We do this by looking at a subsample of 156 stocks that traded before and after the implementation of the order handling rules. The period used as the pre-compliance period is the 10 trading days from 6 January 1997 to 17 January 1997, and the post-compliance period is the last 10 trading days in our sample for each stock (17 September 1997 to 30 September 1997).

Table 10 shows that spreads fell sharply after the order handling rules were implemented, dropping from 3.61% to 2.96%. A natural consequence of narrowing the spread is that fewer trades occur within the spread. Indeed, in the post compliance period, only 14.29% of trades occur inside the spread (compared to 20.44% prior to the implementation of the order handling

[22] We thank the referee for pointing out this point to us.

rules). After the order handling rules, the classification success is lower within the spread and at the midpoint. Classification success decreased at the quotes (the difference is significant at the one percent level) and outside the quotes (significant at the ten percent level). Overall the success rate of the tick-quote classification rule is somewhat lower after the implementation of the order handling rules.

6. Summary

We have documented the success of various trade classification algorithms on a sample of Nasdaq data. Overall, we find that extant trade classification algorithms perform adequately in sorting buy and sell trades, with approximately 81% of all trades being correctly classified by the Lee-Ready algorithm. The results also indicate that despite the median delay of 15 seconds in Nasdaq trades, incorporating those delays into the trade direction algorithm does not improve its accuracy. For researchers using publicly available data sets (such as TAQ) to investigate Nasdaq trades, we recommend using a simple algorithm relying more on tick rules than quote rules. This improves the accuracy of classifying trades in general, and it is substantially more accurate when classifying trades for the purpose of calculating effective spreads.

While our results suggest a reasonable level of accuracy in classifying trade direction, we raise two cautionary issues for consideration by empirical researchers. A sobering aspect of our findings is the difficulty of classifying trades both inside of and outside of the spread. Trades outside of the spread are more likely in times of greater trade intensity, but these times also correspond to the events (such as earnings announcement, takeovers, equity issuance, etc.) that are of greatest interest to researchers. Consequently, trade classification errors may arise non-uniformly over a sample period.

Although this outside-the-spread problem appears specific to Nasdaq data, another difficulty may be more universal. It is generally true that trades at prices other than quotes are becoming increasingly common on all markets. ECN trades are a case in point. We found that only 32% of ECN trades take place at the quotes as compared to over 72% of customer trades on Nasdaq. For our sample, ECN trades are only a small fraction of trades, but, for Nasdaq as a whole, recent data show that ECNs account for more than 30% of trades.[23]

---

[23] See, for example, "How Electronic Networks Snag Trades", Wall Street Journal, March 1, 1999.

We did find that the tick rule performed reliably better than the quote rule for trades away from the quotes, and using the tick rule for all trades away from the quotes, and the quote rule for trades at the quotes provides an improvement over the Lee-Ready algorithm. This improvement arises primarily because the new algorithm is better at classifying trades in between quotes. However, trades between the quotes only accounted for 12.7% of our test sample. For the trades that we excluded from our test sample (24% of all trades), the occurrence of trades within the quotes was much higher (23.2%) suggesting that overall, using the tick-quote rule rather than the Lee-Ready algorithm may be more reliable for Nasdaq data.

These findings suggest that empirical researchers will face increasing difficulty in accurately classifying trade direction. One impact of this is the difficulty in accurately calculating effective spreads, a common measure of transaction costs. We showed that when using our trade classification rule, the average effective spread is much lower than the effective spread calculated using the Lee-Ready algorithm. However, all classification algorithms overstate the true effective spread, suggesting that researchers may generally be overstating the estimated cost of trading Nasdaq stocks.

As trading fragments, it will be more difficult to infer the underlying order, particularly if trades take place inside the spread. The development of more complete data sources may alleviate this difficulty, as may changes in market reporting rules.[24] But it is unlikely that such complete data or order transparency will be available for all markets. Consequently, research on trading algorithms takes on a rising importance for empirical research.

[24] Changes to trade reporting on Nasdaq, including the move to include whether the trade came from a limit or market order (NASD Rules 6950 – 6957 the Order Audit Trail System (OATS) Rules adopted March 1998) will allow better determination of the origin of trades.

References

Aitken, M. and A. Frino, “The accuracy of the tick test: Evidence from the Australian stock exchange”, *Journal of Banking and Finance* 20, (1996) 1715-1729.

Barclay, M., W. Christie, J. Harris and E. Kandel, “The Effects of Market Reform on the Trading Costs and Depths of Nasdaq Stocks”, *Journal of Finance* 54, (1999), 1-34.

Bernhardt, D., Dvoracek, V., Hughson, E. and I. Werner, “Why do Large Orders Receive Discounts on the London Stock Exchange? *University of Illinois Working Paper* 99-0114, (1999).

Blume, M. and M. Goldstein, “Quotes, Order Flow, and Price Discovery”, *Journal of Finance* 52, (1997), 221-244.

Blume, M., A. Mackinlay and B. Terker, “Order Imbalances and stock price movements on October 19 and 20, 1987”, *Journal of Finance* 44, (1989), 827-848.

Ellis, K., R. Michaely and M. O’Hara, “ When the Underwriter is the Market Marker: An examination of Trading in the IPO Aftermarket”, *Journal of Finance*, (2000), *forthcoming*.

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

Hasbrouck, J., “Trades, Quotes, Inventories, and Information”, *Journal of Financial Economics* 22, (1988), 229-252.

Huang, R. and H. Stoll, “Dealer versus auction markets: A paired comparison of execution costs on Nasdaq and the NYSE”, *Journal of Financial Economics* 41, (1996), 313-357.

Lee, C. and M. Ready, “Inferring Trade Direction from Intraday Data”, *Journal of Finance* 46, (1991), 733-746.

Lee, C. and B. Radhakrishna, “Inferring Investor Behavior: Evidence from TORQ Data”, working paper University of Michigan, (1996).

Odders-White, E., “On the Occurrence and Consequences of Inaccurate Trade Classification”, working paper University of Wisconsin, (1999).

Petersen, M. and D. Fialkowski, “Posted versus effective spreads: Good prices or bad quotes?” *Journal of Financial Economics* 35, (1994), 269-292.

Porter, D. and D. Weaver, 1998, “Post-trade transparency on Nasdaq’s national market system,” *Journal of Financial Economics*, Vol 50, no. 2, (1998), 231-252.

Smith, J., J. Selway, L. Reilly and T, McCormick, “NASD Data Relating to the Nasdaq Stock Market, Inc. and its Listed Companies”, working paper NASD (1997).

Table 1
Distribution of Trading Prices and Trading Parties

This table shows the distribution of trades relative to the current quotes, and by trading party. The sample of 2,433,019 trades is all trades time-stamped between 9:30am and 4:00pm on the NASD audit trail. The quotes are the inside bid and ask revisions also taken from the NASD audit trail. Trades take place at the quotes (the bid or the ask), outside the quotes (above the ask or below the bid), at the midpoint quote (the average of the bid and ask price) or inside the quotes but not at the midpoint. Unclassified trades are trades that occur at the beginning of the day before the first quotes. The trading parties are identified by four letter identifiers and come in three categories: market makers who set quotes in the stock, brokers who are NASD members but do not set quotes, and customers who are non-NASD members. Trades are split into three groups: market maker – customer or broker – customer trades, market maker – market maker or market maker – broker trades, and trades via an ECN. Transaction and quote times come from the NASD audit trail.

Trading Price	Cust - Broker	Cust- Mmkr	Mmkr – Broker	Broker – Broker	Mmkr – Mmkr	ECN	Total
Outside Quotes	815	31051	58572	6577	17401	3017	117433
	5.87%	4.99%	4.89%	5.95%	4.32%	3.51%	4.83%
At Quotes	10248	451294	907565	79585	224953	26884	1700529
	73.83%	72.47%	75.81%	72.03%	55.85%	31.25%	69.89%
Inside Quotes	2014	88666	141453	15497	93470	29778	370878
	14.51%	14.24%	11.82%	14.03%	23.21%	34.62%	15.24%
At Midpoint	795	51379	88869	8771	66519	26244	242577
	5.73%	8.25%	7.42%	7.94%	16.51%	30.51%	9.97%
Unclassified	9	366	633	59	438	97	1602
(No Quote)	0.06%	0.06%	0.05%	0.06%	0.11%	0.11%	0.07%
Total	13881	622756	1197092	110489	402781	86020	2433019
	0.57%	25.6%	49.2%	4.54%	16.55%	3.54%	100%

Table 2: Success Rates for Different Trade Classification Algorithms

This table compares different trade classification algorithms. The sample of trades is 1,833,729 classifiable trades, which are trades between a market maker and a customer, a market maker and a broker, or a broker and a customer. The three trade classification algorithms are the quote rule, the tick rule, and the Lee-Ready rule. For each algorithm we give the frequency of trades in each category and the percent that agree with the true trade classification. The categories are total (all trades), and four price categories. "Midpoint" trades occur at the average of the current best bid and best ask quote prices. "Inside spread" trades occur within the best bid and best ask, but not at the midpoint. "At Quotes" trades occur at the bid or the ask, and "Outside Quotes" trades occur below the bid or above the ask.

			Price Range			
	N=1,833,729	Total	Midpoint	Inside Spread	At Quotes	Outside Quotes
Quote Rule	Frequency	92.25		12.66	74.66	4.93
	Percent Agree	76.40		55.23	88.68	64.77
Tick Rule	Frequency	97.80	7.69	12.66	74.66	4.93
	Percent Agree	77.66	60.52	61.32	83.04	65.82
Lee-Ready	Frequency	99.95	7.69	12.66	74.66	4.93
	Percent Agree	81.05	60.52	55.23	88.68	64.77

Table 3: Classification Success by trade size

This table shows the variation in success rate of the Lee-Ready algorithm across different trade sizes. The trades are grouped by the number of shares in the trade, and for each size category the total success, as well as the success at each price category is reported. The price categories group the trades based on the trade price relative to the current best quotes. The distribution of the trades is given (frequency in each category) and the percentage of trades for which the Lee-Ready classification agrees with the true trade classification. The sample size is 1,833,729 trades.

Trade Size	Total	Midpoint	Inside Spread	At Quotes	Outside Quotes
			Frequency		
< 200 shs	31.44	5.46	9.30	80.31	4.94
< 500 shs	24.78	6.81	11.28	77.77	4.13
< 1,000 shs	20.85	8.15	13.07	74.36	4.42
< 5,000 shs	17.39	10.58	17.13	66.68	5.61
< 10,000 shs	3.20	13.14	21.58	58.19	7.10
> 10,000 shs	2.29	14.29	23.85	51.70	10.16
			Percent Agree		
< 200 shs	81.73	57.16	54.32	87.68	63.74
< 500 shs	81.38	56.84	55.87	88.01	66.80
< 1,000 shs	80.70	57.77	55.44	88.62	64.39
< 5,000 shs	80.61	64.21	56.13	91.05	62.17
< 10,000 shs	80.26	72.74	54.57	92.96	68.17
> 10,000 shs	77.85	74.92	51.81	91.87	71.76

Table 4: Classification success by rapidity of trading and quoting

This table examines the relationship between classification success and the speed of trading. Panel A groups trades by the number of seconds between the previous trade and the current trade, as well as the current trade price relative to the quote prices. Panel B groups trades by the time between current trade and previous quote update, and price of current trade relative to quote prices. The frequency of each time delay group is presented, as well as the success rate of the Lee-Ready algorithm for each group. The sample size is 1,833,729 trades.

Panel A: Time from Previous Trade

	Total	Midpoint	Inside Quotes	At Quotes	Outside Quotes
			Frequency		
< 5 seconds	25.94	8.39	12.67	69.56	9.26
5-90 seconds	37.83	8.66	13.80	72.48	5.03
90 – 300 seconds	13.82	6.82	12.00	78.72	2.44
> 300 seconds	22.40	5.78	11.13	81.76	1.30
			Percent Agree		
< 5 seconds	72.58	59.69	50.92	79.50	62.97
5-90 seconds	80.43	62.53	55.19	88.46	65.25
90 – 300 seconds	86.15	60.86	57.64	93.29	67.54
> 300 seconds	88.77	56.58	59.41	95.32	73.24

Panel B: Time from Previous Quote

	Total	Midpoint	Inside Quotes	At Quotes	Outside Quotes
			Frequency		
< 5 seconds	6.02	6.98	13.07	67.97	11.91
5-90 seconds	23.91	7.51	12.35	70.24	9.84
90 – 300 seconds	14.81	7.12	11.03	76.69	5.12
> 300 seconds	55.25	8.00	13.19	76.76	1.99
			Percent Agree		
< 5 seconds	78.36	58.08	58.46	86.76	64.63
5-90 seconds	75.24	57.88	56.31	81.99	64.41
90 – 300 seconds	81.38	59.75	54.58	88.61	61.53
> 300 seconds	83.77	62.00	54.60	91.53	67.85

Table 5: Logistic Regression of Classification Success

The table gives the parameter estimates, standard errors and odds ratios for the following logistic regression:

Log(Pr(correct)/Pr(incorrect)) = a + b1 * outside + b2 * midpoint + b3 * quotes + b4 * trade size + b5 * firm size + b6 * trading speed + b7 * quote speed.

The dependent variable is the logistic transformation of a binary variable that is 1 if the classification via the Lee-Ready algorithm is correct, and 0 if incorrect. Outside, midpoint, quote are dummy variables that take the value 1 if the trade occurs at the named range, and zero otherwise. Trade size is the size of the trade in thousands of shares. Firm size is the average market capitalization of the firm during the trading period in tens of millions of dollars. Trade (quote) speed is the number of minutes since the previous trade (quote). The Slope is the marginal effect of a change in the independent variable on the probability of correct classification, holding all other variables at their mean values. Half of the trade sample was used to estimate the regression: the sample size is 916,864 trades.

Independent Variable	Regression Coefficient	Standard Error	Slope
Intercept	0.0692*	0.00687	
Outside	0.4876*	0.0123	0.0718
Quote	1.8606*	0.00743	0.2740
Midpoint	0.2745*	0.0102	0.0404
Trade Size	0.00554*	0.000521	0.0008
Firm Size	-0.00025*	0.000025	-0.00004
Trade Speed	0.00926*	0.000225	0.0014
Quote Speed	0.000749*	0.000043	0.0001
2 log L	90974*		

* indicates that the Chi-square statistic of the estimate has a p value of 0.01.

Table 6: Classification success for Tick-Quote Rule

This table compares the success of the Lee-Ready algorithm with a proposed alternative (Tick-Quote rule). The alternative algorithm uses the quote rule to classify trades at the quote (bid or ask) and the tick rule to classify all other trades. The distribution of trades across the price categories as well as the success rates of the Lee-Ready and proposed algorithms is given.

Trading Price	Frequency	Lee-Ready Percent Agree	Tick – Quote Rule Percent Agree
Outside Quotes	4.93	64.77	65.82
At Quotes	74.66	88.68	88.68
Inside Spread (not at midpoint)	12.66	55.23	61.32
Midpoint	7.69	60.52	60.52
Unclassified	0.05		
Total	1,833,729	81.05	81.87

Table 7: Trade Classification Success: Excluding the First Month of Trading and Market Maker – Broker Trades

Panel A shows the success rate of the tick-quote rule for two different sub-samples: customer trades excluding the first month of trading, and market maker – broker trades. The tick-quote rule classifies trades at the bid (ask) as buys (sells) and all other trades by the tick rule. The distribution of the trades with respect to the quote prices is given, as well as the success of the tick – quote rule in each range. Customer trades are all trades between a market maker and customer, or broker and customer that occurred after the first 20 days of trading.

Panel B shows the parameters from a logistic regression where a binary variable denoting correct or incorrect classification is regressed against dummy variables for the trading range (outside, midpoint, at quotes), trade size, trading speed, quoting speed, firm size. The parameters and their marginal effects on the probability of correct classification (slope) are reported.

Panel A: Distribution of Classification Success

	Total	Price Range			
		Midpoint	Inside Spread	At Quotes	Outside Quotes
Excluding the First Month of Trading					
Frequency	1,380,271	7.40	13.37	74.29	4.89
Lee-Ready % Agree	79.80	60.56	54.71	87.41	62.81
Tick-Quote % Agree	80.80	60.56	61.55	87.41	64.48
Market Maker – Customer Trades					
Frequency	636,637	8.20	14.24	72.50	5.01
Lee-Ready % Agree	83.01	64.83	57.96	91.19	66.56
Tick-Quote % Agree	84.24	64.83	66.15	91.19	67.90

Panel B: Logistic Regression

Independent Variable	Excluding First Month of Trading		Market Maker – Customer Trades	
	Estimated Regression Coefficient	Slope	Estimated Regression Coefficient	Slope
Intercept	0.0473*		0.2417*	
Outside	0.4193*	0.0645	0.4278*	0.0613
Quote	1.7278*	0.2657	2.0095*	0.2878
Midpoint	0.2656*	0.0408	0.3147*	0.0451
Trade Size	0.0087*	0.0013	0.0005	0.0001
Firm Size	-0.0001*	-0.00002	-0.0005*	-0.0001
Trade Speed	0.0105*	0.0016	0.0061*	0.0009
Quote Speed	0.0008*	0.0001	0.0006*	0.0001
2 log L	121316*		66002*	

* signifies the parameter is significant at the 1% level, standard errors are not reported.

Table 8: Reporting delays on the TAQ database

This table examines the reporting delays for trades and quotes on the NYSE TAQ database. Trades are matched for each day using the trading price, volume, and time proximity if there are multiple trades at the same price and for the same volume. The reporting delay is the time between the TAQ timestamp and the NASD audit trail timestamp. As well as the frequency of each time delay, the table reports the occurrence of a quote update between the NASD time and the TAQ time. Trades with a quote update between the true time and the reported time will use the wrong quote in the trade classification algorithm. Panel B shows the difference in the timestamp for inside quote revisions on TAQ and the NASD audit trail.

Panel A: Time delay for Trades

Time Delay	Frequency	Percent	# with quote update in between	# with quote update in between (as % of trades with this delay)	# with quote update in between (as % of all trades)
-10 minutes	1377	0.06	917	66.59	0.04
-1 minute	1061	0.05	404	38.08	0.02
<-1 minute	6569	0.28	517	7.87	0.02
Same time	61088	2.59	0	0	0
1 second	292260	12.41	14418	4.93	0.61
2 seconds	146342	6.21	11641	7.95	0.49
3–14 seconds	315858	13.41	44363	14.05	1.88
15 seconds	743896	31.58	81967	11.02	3.48
16 seconds	636182	27.01	76778	12.07	3.26
17 – 60 seconds	133335	5.66	20398	15.3	0.87
1 - 10 minutes	15896	0.67	5035	31.67	0.21
More	1720	0.07	1078	62.67	0.05
Total	2355584	100	257516		10.93

Panel B: Time Delay for Quotes

Time Delay	Frequency	Percent
More than -10 minutes	420	0.1
-10 minutes to –31 seconds	167	0.0
-5 to -30 seconds	120	0.0
-2 to -4 seconds	14061	2.4
-1 second	69990	11.8
Same time	391431	66.0
1 second	60635	10.2
2-4 seconds	51259	8.6
5–30 seconds	4021	0.7
31 seconds to 10 minutes	1038	0.2
More	202	0.0
Total	593344	100.0

Table 9: Success of Trade Classification with Reporting Delays.

This table compares the success of the Tick-Quote trade classification algorithm using the NASD database (true trade time), and the TAQ database (TAQ trade time), plus the TAQ trade time minus 15 seconds (the median reporting delay). The number of trades is 1,751,792 that are the trades that we could directly sign and that had a matching trade on the TAQ database.

N=1,751,792	True Trade Time	TAQ Trade Time	TAQ Trade Time – 15 seconds
Lee-Ready % Agree	80.19%	81.95%	80.77%
Tick-Quote % Agree	82.74%	83.74%	82.73%

Table 10: Estimate of the Impact of 1997 Order Handling Rules on Trade Classification

This table compares two samples: market maker – customer trades from September 17 to September 30, 1997, and market maker – customer trades from January 6 to January 17, 1997. Only stocks that traded in both periods are included (156 stocks). The first sample proxies for trades after the introduction of the order handling rules, and the second sample is before the implementation of the order handling rules. We do not know the exact date of implementation for each stock. The table gives the daily average time weighted percentage spread for each period, and the proportion of trades that occur at the midpoint, within the spread, at the quotes and outside the quotes, and the success of the tick-quote rule in classifying the direction of these trades. A t-test was performed to test the difference in spreads, and Pearson's Chi-square test was done to test for differences in the success rate pre- and post- order handling rules.

	Average Percentage Spread	Trades	Total	Price Range			
				Midpoint	Inside Spread	At Quotes	Outside Quotes
Pre Order Handling Rules	3.61	Frequency	39821	7.57	12.87	77.13	2.36
		Tick-Quote % Agree	86.23%	63.03	64.90	92.65	69.47
		Lee-Ready % Agree	84.86%	63.03	54.66	92.65	67.45
Post Order Handling Rules	2.96	Frequency	53946	5.79	8.50	79.70	5.97
		Tick-Quote % Agree	76.67%	58.41	60.79	80.65	64.54
		Lee-Ready % Agree	76.16%	58.41	55.06	80.65	64.11