Inter-Temporal Trade Clustering and Two-Sided Markets

Asani Sarkar
Senior Economist
Federal Reserve Bank of New York

Robert A. Schwartz
Professor of Finance
Zicklin School of Business
Baruch College, CUNY

Avner Wolf
Professor of Finance
Zicklin School of Business
Baruch College, CUNY

Current Version: November 22 2005

We thank Markus Brunnermeier, Michael Goldstein, Joel Hasbrouck, Milt Harris, Terry Hendershott, Murali Jagannathan, Charles Jones, Eugene Kandel, Bruce Lehmann, Albert Menkveld, Lasse Pederson, Ioanid Rosu, Krystin Ryqvist, Gideon Saar, George Sofianos, Shane Underwood, Jiang Wang, and James Weston, and seminar participants at the AFA 2006 meetings, the Federal Reserve Bank of New York, the NBER Market Microstructure conference of October 2005, the Microstructure conference in Norges Bank (Oslo), Baruch College, Rice University, Rutgers University, and SUNY Binghamton for helpful comments. The views stated here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of New York, or the Federal Reserve System.

Abstract

Inter-Temporal Trade Clustering and Two-Sided Markets

We show that equity markets are typically two-sided and that trades cluster in certain trading intervals for both NYSE and Nasdaq stocks under a broad range of conditions – news and non-news days, different times of the day, and a spectrum of trade sizes. By "two-sided" we mean that the arrivals of buyer-initiated and seller-initiated trades are positively correlated; by "trade clustering" we mean that trades tend to bunch together in certain intervals with greater frequency than would be expected if their arrival was a random process. Controlling for order imbalance, number of trades, news, and other microstructure effects, we find that two-sided clustering is associated with higher volatility but lower trading costs. Our analysis has implications for trader behavior, market structure, and the process by which new information is incorporated into market prices.

Inter-Temporal Trade Clustering and Two-Sided Markets

In this paper, we characterize the joint arrivals of buyer-initiated and seller-initiated trades in intervals of half-hour to a minute, and examine its association with price volatility and trading costs, for a sample of NYSE and Nasdaq stocks. We call intervals with unusually many buyer-initiated trades and unusually few seller-initiated trades, or vice versa, as one-sided. Conversely, intervals with unusually many (few) buyer-initiated trades *and* unusually many (few) seller-initiated trades are two-sided. In addition, we examine whether trades cluster only on one side of the market or on both sides together. By clustering, we mean that periods with unusually high numbers of trades are more prevalent than would be expected under random trade arrival. Our primary motivation is that sidedness and clustering reflect the trading motives of participants and sheds light on the price formation process.

Broadly speaking, the literature emphasizes three motives for trading, each of which has different implications for sidedness and clustering, and their association with trading costs and volatility. Trading based on superior information predicts trade clustering on one side of the market, with high volatility and trading costs during these periods. The second motive, that trading is based on differential information or on divergent beliefs, leads to trade clustering together on both sides of the market, with elevated volatility during these periods. Finally, trades to rebalance portfolios may also imply two-sided markets, but do not imply a systematic relation between sidedness, volatility, and trading costs. In the next section, we discuss in greater detail the different trading motives, the associated literature, and to what extent our findings support these motives.

We find that, for practically every stock in our sample, the arrivals of buyer-initiated and seller-initiated trades within time intervals are positively correlated (i.e., that buyers trigger more trades in an interval when a larger number of sellers have triggered trades in that same interval, and vice versa). In other words, markets are *two-sided*. Further, buyer-initiated and seller-initiated trades tend to cluster together in particular intervals. We find that two-sided trade clustering occurs under a broad range of conditions – news and non-news days, different times of the day, alternative market structures, and a spectrum of trade sizes. These findings underscore the importance of divergent beliefs and differential information as motives for trading.[1]

However, the validity of one motive need not be at the expense of another, and all three motives discussed earlier are likely to be important to varying degrees in understanding the dynamics of trading in financial markets. In particular, if information is short-lived, trading based on asymmetric information may be observable over short windows. Indeed, we find that for NYSE stocks and for 1-minute intervals, one-sided trading is evident in the first 15 minutes of days with news. However, for windows longer than a minute and for Nasdaq stocks, two-sided trading remains the norm. Thus, trading motivated by asymmetric information appears to play a role over short windows on days with news.

Next, we use regression analysis to examine the association between sidedness, price volatility (measured by high-low price ranges) and trading costs (measured by the bid-ask spread), after controlling for order imbalance, number of trades, time-of-day effects, news arrival and the share price. We find that the sidedness and clustering are significantly correlated with volatility and trading costs. Specifically, volatility is highest in intervals with two-sided clustering (i.e. with large numbers of buyer-initiated and seller-initiated trades), after accounting for trading costs; and trading costs are highest when markets are one-sided (i.e. many buyer-initiated but few seller-initiated trades, or vice versa), after accounting for volatility. These findings obtain for interval lengths of 30 minutes to one minute. This result is consistent with models where trading is motivated by divergent beliefs or differential information.

We conduct a number of additional investigations to determine the robustness of the results. We examine the effect of inaccuracies in classifying the trade direction by examining trades executed inside quotes and at mid-quotes, and trades of large stocks---instances where trade classification algorithms are less accurate than for trades occurring at the quote (Ellis, Michaely and O'Hara, 2000; Peterson and Sirri, 2003). Since our sample is for the post-decimalization period, when trade sizes decreased and large orders were more likely to have been broken up, we further address the generality of our results by repeating the analysis for a pre-decimalization period. Next, we examine alternative methodologies for estimating the

[1] Recent work underscores the importance of heterogeneous beliefs. Bamber et al (1999) provides evidence that differential interpretations are an important stimulus for speculative trading. In the context of asset pricing, Anderson et al (2004) show that dispersion of earnings forecasts is a priced factor in traditional factor asset pricing models and is a good predictor of return volatility in out-of-sample tests.

correlation of buyer-initiated and seller-initiated trades and alternative measures for news arrival and for volatility. In all cases, we show that markets exhibit two-sided clustering.

In a related paper, Hall and Hautsch (2004) use limit order data for 3 actively-traded stocks on the Australian Stock Exchange, and find that buy and sell intensities evaluated at the time of each transaction are strongly positively auto-correlated and cross-correlated. Using a different methodology, we find the same result for a large number of stocks, both active and less active. Different from Hall and Hautsch (2004), we discern variations in the degree of sidedness and clustering (e.g. extreme versus moderate one-sidedness), in addition to just a correlation, and relate this variation to trading costs and volatility.

In a series of papers, Easley, Kiefer and O'Hara (1996, 1997a, 1997b) explore the roles of the direction and sequence of trades in the price formation process. More buys (sells) are expected on days with good (bad) events, and fewer trades arrive on days with no information events. The authors estimate the arrival rates of informed and uninformed traders using the structure of an asymmetric information microstructure model, and data on the daily numbers of buyer-initiated and seller-initiated trades, and the number of no-trade outcomes. Easley, Engle, O'Hara, and Wu (2005) allow the arrival rates of informed and uninformed traders to vary from one day to the next. In their model, the absolute trade imbalance contains information on the arrival of informed trades, while the balanced trade (i.e. the difference between total trades and the absolute imbalance) contains information on the uninformed trades.

We find that markets are *more* one-sided and *less* two-sided in periods of high imbalance (defined as the log ratio of the absolute imbalance to total trades) while the reverse is true in periods with many trades. These results are consistent with Easley et al (2005) who find that an increase in the share of imbalanced trades forecasts higher arrival rates of informed traders (likely leading to more one-sided markets). However, we also find that, even during periods with high imbalance and few trades, markets are two-sided. This suggests that our sidedness variable contains information not fully captured by order imbalance and total trades. In particular, the sidedness of markets depends on the relation between the distributions of buyer-initiated and seller-initiated trades, whereas order imbalance is a summary measure of these distributions (i.e. the difference between the total numbers of buy and sell trades).

Prior research relates the buy-sell imbalance to liquidity and volatility. Hall and Hautsch (2004) find that the instantaneous buy-sell imbalance is a significant predictor of returns and volatility. Chordia, Roll and Subrahmanyam (2002) show that daily order imbalances are negatively correlated with liquidity. We find that trading frequency is positively related to volatility and negatively to trading costs, consistent with past research, whereas the imbalance variable yields conflicting and statistically less significant results. Even after controlling for imbalance and total trades, however, we find that our sidedness variables are highly significant in explaining volatility and trading costs. This result further demonstrates that sidedness is informative even after accounting for the buy-sell imbalance.

Our analysis is related to the literature that uses the autoregressive conditional duration (ACD) method to model inter-trade arrival times (e.g. Dufour and Engle, 2000). While ACD models focus on the time required for the market to absorb a given amount of volume, our analysis considers the trading intensity in a specified unit of time. When we examine the intensity of trade arrivals (independent of whether the trade is buyer-initiated or seller-initiated), we find that trades cluster in certain intervals and that volatility and trading costs tend to be highest in these periods. These results are consistent with those of Engle and Russell (1994) who find evidence of co-movements among duration, volatility, volume, and spread, and with Engle (1996) who finds that shorter durations lead to higher volatility. In contrast with the ACD models, we also examine the cross-correlation between the arrivals of buyer-initiated and seller-initiated trades and, as noted, we consider the implications for volatility and trading costs.

Our methodology is different from, and complementary to, those in Hall and Hautsch (2004), who use a bivariate dynamic intensity model, or Easley et al (2005) who use a maximum likelihood method combined with a GARCH-type process for forecasting trade arrivals, or the ACD models. These methodologies are stock-specific and computation-intensive, allowing for relatively small numbers of actively traded stocks to be analyzed; thus, no conclusions can be derived about the market (as noted by Easley et al, 1997a).[2] We, in contrast, aggregate across relatively large numbers of stocks, both active and less active, and compare aggregate trade clustering for different conditions (e.g. between days with and without news). Whereas our approach is essentially static, an advantage of the other methodologies is that they incorporate the dynamics of the buy and sell arrival processes; further, they allow for an interaction between

[2] Hall and Hautsch (2004) examine 3 stocks, and their results are robust only for the most active stocks. Easley et al (2005) choose high volume NYSE stocks and exclude days with either no buys or no sells. Easley et al (1996) aggregate across stocks after assuming that the information content of each stock is the same, thus reducing the number of parameters that need to be estimated.

the dynamics and price formation.[3] Further, the Hall and Hautsch (2004) and the ACD methods require no aggregation over time. In general, we view our methodology as providing an alternative way of analyzing the process of trade arrivals via our sidedness and clustering variables. In this context, it is reassuring that some of our results are consistent with those found using these alternative methodologies.

Our paper is organized as follows. In Section 1, we discuss alternative models of trading motives and price formation, and how our results relate to predictions from these models. In Section 2, we describe our data and present descriptive statistics. In Section 3, we examine the joint distribution of buyer-initiated and seller-initiated trades (i.e. the sidedness of markets) for NYSE and Nasdaq stocks. In Sections 4 and 5, we assess the relationship between trade clustering, sidedness and, respectively, price volatility and trading costs. In section 6, we examine whether the results are sensitive to errors in classifying the trade direction, to the sample period (i.e. pre- or post-decimalization), to different lengths of the trading interval, to alternative measures of news and volatility, and alternative methodologies for estimating sidedness. We conclude in Section 7 by considering the broader implications of our study. In Appendix A, we provide details of our methodology for estimating the joint distribution of buyer-initiated and seller-initiated trade arrivals. In Appendix B, we discuss results on clustering of aggregate trades (independent of whether they are buyer or seller-initiated).

1. Trading Motives and Price Formation: Alternative Views

The role of trading activity in price formation is a central topic in market microstructure. A key idea in the literature is that investors' trading motives may be inferred from trading activity, including the number and sign of trades, trade size and the duration between trades. For example, a preponderance of buy (sell) orders may signal good (bad) news, causing traders to revise upward their expected value for the stock (Easley et al, 1997a). Hasbrouck (1991) shows empirically that the market maker, from observing trade attributes such as sign and size, infers information from the trade sequence.

Trading may occur due to asymmetric information (i.e. some investors have superior information to others); differential information (i.e. some investors have different information

[3] For example, in ACD models, the current duration can depend on past durations, and the duration simultaneously affects quote revisions and the correlation between current and past trade direction.

than others) or heterogeneous beliefs (i.e. investors have different interpretations of news); and portfolio rebalancing. Alternative trading motives have distinct implications for the sidedness and clustering of trades, and the relationship between sidedness, volatility and trading costs. Table 1 provides a summary of the implications for the three types of investor motives. Since our focus is on trade arrivals and clustering, we mostly limit discussion to dynamic, rather than static, trading models.

When some investors have superior private information (Model 1), a one-sided market is likely to occur (Wang, 1994; Llorente et al 2002).[4] If, for example, the informed trader sells the stock upon receiving a bad signal, the price decreases in the current period. Since the private information in only partially revealed in the price (when the equilibrium is not fully revealing), the insider is likely to sell again in the next period.[5] Dufour and Engle (2000) suggest that insiders may trade quickly to prevent information leakage, implying that trades are likely to be clustered on one side of the market following news events.[6] Asymmetric information among investors can cause price volatility to increase (Wang, 1993, 1994) because less-informed investors demand additional risk premium as compensation for the risk of trading against better-informed traders. This results in increasing price elasticity to supply shocks and higher price volatility.[7] Higher volatility and more adverse selection imply that trading costs are also higher with asymmetric information. Also, since asymmetric information leads to one-sided markets, dealer's inventory imbalance is likely to be greater and further increase trading costs.

When investors observe different information signals (Model 2A), they may buy or sell the stock depending on their particular information signal, implying that informed trading can be

---

[4] We note that one-sided order flow would not obtain in models where price changes follow a martingale (e.g. Kyle, 1985) since if the price change is proportional to order flow (with a fixed constant of proportionality), then order flow must also be a martingale.
[5] This assumes that the informed trader places market orders. If, instead, he places aggressive limit sell orders on receiving a bad signal, then we may observe a sequence of buyer-initiated trades as market orders from the opposite side hit the informed trader's limit orders. Even in this case, however, a one-sided market obtains.
[6] Numerical solutions in Foster and Viswanathan (1994) also suggest the possibility of trade clustering in early and late periods after the arrival of information.
[7] According to Wang (1993), volatility may decrease with asymmetric information because uninformed investors have better information about the fundamental value of the stock (due to the information from insider demands and prices) which reduces the uncertainty in future cash flows. However, if there is enough adverse selection in the market, the net effect is for volatility to increase.

observed on both sides of the market.[8] Investors trade many rounds in the equilibrium where prices are not fully revealing; further, clustering can occur as investors maintain aggressive speculative positions at early dates (He and Wang, 1995). Differential information is associated with higher volatility since dispersion magnifies the effect of noisy information on price volatility (Grundy and McNichols, 1989; Shalen, 1993). The effect of differential information on trading costs is unclear. Uncertainty in the value of the stock tends to decrease liquidity (He and Wang, 1995). On the other hand, dealers and limit order traders face lower risk from unbalanced inventory or portfolio positions in two-sided markets, which increases liquidity.

Investors may interpret a public signal differently (Model 2B), with implications for sidedness, trading costs and volatility that are similar to Model 2A. We expect that trading based on differential interpretations to lead to a two-sided market. For example, in Kandel and Pearson (1995), trade occurs because agents use different likelihood functions to interpret public news. Trades may be two-sided because one agent can interpret the public signal more optimistically or pessimistically than the other.[9] While models of differential interpretation do not by themselves predict clustering, clustering is likely if, as trades occur, further trades are executed due to order flow externalities (i.e. orders attracting orders).[10] In Kim and Verrecchia (1994), some traders process public news into private, and possibly, diverse information about a firm's performance; the information can be interpreted as informed judgments or opinions. They show that as the diversity of information increases, there are more information processors, leading to higher volatility and trading costs. As in Model 2A, the overall relation of diverse opinions with trading costs is ambiguous if trading costs are lower in two-sided markets due to the inventory effect.

Finally, investors may trade to rebalance their portfolios (Model 3). If returns of traded and non-traded assets are correlated, then uninformed investors sometimes buy and at other times sell in order to hedge their non-traded risk, leading to two-sided markets (Wang, 1994; Llorente et al 2002). But, since there is no large change in expectations or uncertainty about stock value,

---

[8] He and Wang (1995) provide an example of two-sided trading purely due to differential information. In the example (footnote 18 in their paper), half of the investors estimate, based on their information, that the supply shock has increased and buy the stock, while the other half estimate that the supply shock has decreased and sell the stock.

[9] Harris and Raviv (1993) develop a model of divergent interpretations where two groups of traders agree whether a signal is positive or negative, but one is more "responsive" to the information. When the cumulative signal is positive (negative), the more responsive (unresponsive) group buys all available shares. As the cumulative signal changes sign, the direction of trades also changes.

[10] Hendershott and Jones (2005) and Antunovich and Sarkar (2005) provide empirical evidence on order flow externality.

rebalancing trades do not generate additional volatility or trading costs (He and Wang, 1995).

How do our findings relate to the various trading motives? Overall, as indicated in Table 1, we find that markets exhibit two-sided clustering and that volatility is highest during such periods. These findings are consistent with the predictions of model 2A (differential information) or model 2B (divergent beliefs).[11] Two-sided markets are also predicted by model 3 (portfolio rebalancing) but our findings on volatility and trading costs are inconsistent with its predictions. In addition, because of the short (30 minutes to 1 minute) assessment intervals used in our empirical analysis, one would expect that the transactions costs involved in such frequent trading would be prohibitive for investors seeking to rebalance their portfolios.[12]

It seems difficult to reconcile the findings of two-sided clustering and the high volatility during such periods with Model 1 (asymmetric information models). Two-sided clustering may occur in such models if discretionary liquidity traders and informed traders cluster in the same period (Admati and Pfleiderer, 1988) or if uninformed trades are serially correlated within a day.[13] However, two-sided clustering obtains even when we separately examine a sample of large trades, which are less likely to be from uninformed traders; and it also obtains immediately following news arrival and for intervals as short as 2-minutes (1-minute for Nasdaq stocks). However, for NYSE stocks and for 1-minute windows, we find evidence of one-sided trading in the first 15 minutes of days with news. We conclude that trading based on asymmetric information appears to be short-lived and accounts for a small amount of trading. This is consistent with results from Easley et al (2005) who find that the probability of informed trading is relatively small, varying between 8% and 19% for the 16 stocks in their sample.

Our findings suggest that having different interpretations of public news and/or different private information signals leads participants to trade on opposite sides of the market. The prevalence of two-sided trading on days characterized by significant news release is particularly striking. While participant responses may not quickly produce a new equilibrium value for a stock, they apparently do move prices rapidly into new ranges within which some participants are buyers, and other are sellers, depending on their individual assessments of the news.

---

[11] Frankel and Froot (1990) also find a positive association between dispersion and price volatility.
[12] Informal conversations with practitioners reveal that index funds, for instance, do not rebalance with undue frequency (e.g., every half hour).
[13] However, Easley et al (2005) find that an increase in the arrival of informed traders forecasts a *decrease* in the arrival rates of uninformed traders; further, uninformed trades are highly persistent across days.

2. Data and Descriptive Statistics

We use time-stamped trade and quote data from the Transactions and Quotes (TAQ) Database of the New York Stock Exchange (NYSE), which records transaction prices and quantities of all trades, as well as all stock quotes that were posted. Our data are for the period January 2 to May 28, 2003, for a matched sample of 41 NYSE stocks and 41 Nasdaq stocks.[14] To purge the data of potential errors, we delete any trades or quotes with:

1. Zero or missing trade price.
2. Bid or ask prices that are missing, negative or unusually small relative to surrounding quotes.
3. Quotes where the change in the bid (ask) quote, from the previous bid (ask) quote, exceeds $10.
4. The quoted bid-ask spread is negative.
5. The proportional quoted bid-ask spread or effective bid-ask spread is in the upper 0.5 percentile of its distribution by stock and time interval.
6. The quoted bid or ask size is negative.
7. Trade or quote prices that are outside regular trading hours.

These filters eliminated approximately 3% of all recorded prices and quotes. After elimination, the NYSE data include 4,877,678 trades and the Nasdaq data include 10,860,576 trades.

Initially, we examine trade arrivals in half-hour intervals. Later, in section 6, we examine shorter intervals up to 1 minute in length. The final sample contains 54,226 half-hour intervals for NYSE stocks and 54,415 half-hour intervals for Nasdaq stocks. We analyze all trades, as well as a sample of large trades, which are more likely to be information-based trades of institutions (Easley et al, 1997a, find that large trades are twice as informative as small trades). We define large trades, for a stock, as those that are in the top decile of the dollar value of trades for that

---

[14] We initially matched 50 Nasdaq stocks with 50 NYSE stocks but had to drop 9 NYSE stocks mostly as they were acquired by or merged with another company. To match based on market value and closing price, we randomly select 41 NYSE stocks that were trading on the last trading day of December 2002, and then select 41 Nasdaq stocks with a market value and closing price that, at that date, were nearest to those of the NYSE stocks. Specifically, for the j^{th} matching variable, let x_j be the data for NYSE stock x, and y_j be the data for Nasdaq firm y, where j=1 (the market value), or 2 (the closing price). The Euclidean distance between NYSE firm x and Nasdaq firm y is:

$$d(x,y) = \sqrt{\sum_{j=1}^{2}\left(x_j - y_j\right)^2} \tag{1}$$

We select a matched Nasdaq firm y to minimize $d(x,y)$. Since variables with large variance tend to have more effect on $d(x,y)$ than those with small variance, we standardize the variables before the minimization.

stock in our sample period. This procedure classifies as large trades those with dollar values that exceed, on average, \$32,665 for NYSE stocks and \$28,251 for Nasdaq stocks.[15]

Table 2 shows basic descriptive statistics for our sample. "All" ("large") refers to all (large) trades. The column labeled "All days" shows results for the entire time period. Panel A is for NYSE stocks and Panel B is for Nasdaq stocks. On December 31, 2002, market capitalization averaged \$4.7 billion for NYSE stocks and \$4.4 billion for Nasdaq stocks, and the closing price averaged \$21.56 for NYSE stocks and \$21.35 for Nasdaq stocks (the respective values are close because the samples are matched).

The table presents measures of volatility and trading costs, as well as the number of buy-triggered and sell-triggered trades, for different times of the day, and for days with and without news. The reported figures have been multiplied by 100. We define two measures of volatility: (1) ACLOP, which is the absolute value of the return from the previous day's close to the current day's opening price, and (2) HILO, which is the log of the ratio of the maximum to the minimum price in a period. Our measures of trading costs are PQBAS, the proportional quoted half-spread and PEBAS, the proportional effective half-spread. PQBAS is the quoted bid-ask spread divided by 2M, where M is the quote mid-point. PEBAS is Q(P- M)/M, where P is the trade price, and Q is +1 (-1) for a buy- (sell-) triggered trade, respectively.

The second and third columns of Table 2 show descriptive statistics for news days and non-news days. To isolate news days, we select the 30 percentile of days where the value of ACLOP is largest (later, in section 6, we directly identify days with firm-specific news events).[16] HILO and PQBAS are significantly higher on news days for both exchanges. PEBAS is higher on news days for NYSE stocks but not for Nasdaq stocks. For both markets, there are more trades, and greater volume as compared to non-news days.

The last five columns of Table 2 show statistics for the first, last and intermediate half-hours. The first and last half-hours are further divided into 15 minutes intervals. Volatility, trading costs and trading activity are all higher in the first half-hour (and the first 15 minutes, in

[15] According to Campbell, Ramadorai and Vuolteenaho (2004), trades that are over \$30,000 in size are highly likely to be initiated by institutions. Our trade size cutoff is a close match to their number, which provides some assurance that our procedure may distinguish between institutional and retail trades. Institutional trading volume accounts for a large fraction of market volume. Of course, the institutions' percent of trades is far less than their percent of shares. This implies that the percentage of large trades that is triggered by institutional orders is particularly large.

particular), relative to the middle half-hours, on both markets. Following the first half-hour, we observe a decline in trading activity, trading costs and volatility. Trading activity picks up again 30 minutes before the close, but volatility and trading costs remain low. In the final 15 minutes, trading activity is highest, and volatility and trading costs increase (especially on the Nasdaq market), although they remain below the levels of the opening 15 minutes of the day.

The last four rows of the table show statistics for large trades. There are, on average, only four to five large trades per half-hour interval for NYSE stocks, and only nine to ten large trades per half-hour interval for Nasdaq stocks.[17]

3. Order Clustering and the Sidedness of Markets

In this section, we investigate the joint arrivals of buy-triggered and sell-triggered trades. We examine whether buyer-initiated and seller-initiated trades are correlated and the extent to which they cluster in particular intervals. Clustering is defined as an unusually high number of trades arriving in particular intervals. Referring to Table 1, if trades are based on asymmetric information, we expect the arrivals to be clustered on one-side of the market. Alternatively, for trading based on differential information or beliefs, we expect the arrivals to be clustered on both sides of the market together. Finally, if trading is mainly due to portfolio rebalancing, markets may be two-sided but without an implication for clustering.

Trade clustering may be explained by market participants in general, and by institutional investors in particular, making strategic timing decisions. Accordingly, we separately study large trades because these trades are more likely to be made by institutions that market time their orders. These trades are also of particular interest because institutions are more apt than retail investors to be informed, and thus their order flow is more likely to be one-sided than the retail order flow. On the other hand, institutional order flow may also be two-sided to the extent that portfolio managers have diverse motives for trading. For instance, some institutional investors are thought to have superior information concerning share value (e.g., the value funds), others look only to passively mimic an index (e.g., the index funds), and yet others seek to exploit short-run trading opportunities (e.g., the hedge funds). Even funds within the same category

---

[16] In Easley et al (2005), the probability that an information event occurs on a particular day is between 0.33 and 0.58 for actively traded stocks. Thus, the 30 percentile cut-off is on the low side of this range, but appears reasonable since we have both active and inactive stocks in our sample.

(e.g., value funds) can be on opposite sides of a market if the portfolio managers interpret information differently (i.e., if they have divergent expectations). Consequently, whether institutional order flow is predominantly one-sided or two-sided is an empirical issue.

Recognizing that trade clustering could also be an artifact of pooling periods with heavy trading volume (e.g., the first fifteen minutes of the trading day) with periods when trading is generally lighter (e.g., in the middle of the day), we examine the pattern separately for the first and the last 15 minutes of the trading day. We also consider the possibility that trade clustering is an artifact of pooling information-rich trading periods with periods where little news has occurred. Thus, we present evidence on trade clustering during the first 15 minute period on news days. Finally, we address the possible effect of market structure by comparing the patterns of trade arrivals for NYSE and Nasdaq stocks.

We describe the methodology for determining sidedness in Section 3*A*. Results for individual stocks are in Section 3*B*, and results for the aggregate of stocks are in Section 3*C*.

A. Methodology

We use the Lee and Ready (1991) algorithm to identify transactions as either buy-triggered or sell- triggered. If the trade price is closer to the most recent ask (bid) price in the same stock, it is a buy (sell) initiated trade. For prices equal to the quote mid-point, trades that take place on an uptick are buys, and trades that take place on a downtick are sells. The Lee-Ready (1991) algorithm cannot classify some trades, in particular those executed at the opening auction of the NYSE, and these are omitted from our sample. In section 6, we examine the effects of trade classification errors on our results.

We tabulate the number of buyer-initiated and seller-initiated trades in each interval of each day, and record the number of intervals for which each specific combination of buyer-initiated and seller-initiated trades (e.g., two buy triggered trades and three sell triggered trades in a window) was observed. The results are recorded in a matrix (BSELL matrix from here on). Our null hypothesis is that buy and the sell arrivals (i.e. the rows and columns of BSELL) are not associated. Given our large sample size, the test statistic should be distributed approximately as

---

[17] In general, the number of trades is greater in the Nasdaq market than on the NYSE. Historically, the difference has been attributed to the greater prevalence of dealer intermediation in Nasdaq trading.

chi-square if the null hypothesis is true. To test the hypothesis, we use the Pearson chi-square statistic Q_P which reflects the observed minus the expected frequencies, as follows:

$$Q_P = \sum_i \sum_j \frac{(n_{ij} - \varepsilon_{ij})^2}{\varepsilon_{ij}}, \tag{1}$$

where for row i and column j, n_{ij} is the observed and ε_{ij} is the expected frequency. Under the null hypothesis of independence, $\varepsilon_{ij} = \frac{n_{i.} n_{.j}}{n}$, where $n_{i.} = \sum_j n_{ij}$ is the sum for row i, $n_{.j} = \sum_i n_{ij}$ is the sum for column j, and $n = \sum_i \sum_j n_{ij}$ is the overall total. Further, Q_P has an asymptotic chi-square distribution under the null with $(R\text{-}1)(C\text{-}1)$ degrees of freedom, where R is the number of rows and C is the number of columns. For large values of Q_P, the null hypothesis is rejected in favor of the alternative hypothesis of dependence between the buy and sell arrivals.

We consolidate the BSELL matrix across stocks to make statements about the aggregate of buy and sell arrivals over the sample. Since trading activity (and, hence, the size of the BSELL matrix) differs across stocks, we standardize each BSELL matrix so that stocks with widely different arrival rates are comparable. To this end, the BSELL matrix is mapped for each stock into a 3-by-3, High-Medium-Low matrix (HML matrix from now on). We assume that buy and sell trades follow a random (Poisson) arrival process, with the Poisson parameter λ_b (for buys) equal to the mean number of large buy trades, and the Poisson parameter λ_s (for sells) equal to the mean number of large sell trades in the sample.[18] The mapping rule is based on λ_b and λ_s. Specifically, an interval with n_b buy trades is mapped into the:

- LOW BUY cell if n_b<=Rounddown(λ_b-$\sqrt{\lambda_b}$)
- HIGH BUY cell if n_b> Roundup (λ_b+$\sqrt{\lambda_b}$)
- MEDIUM BUY cell in all other cases.

Intervals with n_s sell trades are similarly mapped into LOW, MEDIUM or HIGH SELL cells based on λ_s. Note that, since λ is a Poisson parameter, $\sqrt{\lambda}$ is the standard deviation of the number of trades for the stock in the sample. Hence, our LOW (HIGH) cutoff represents the mean minus (plus) the standard deviation of the stock's trading frequency. The values of λ_b and

[18] As will be shown below, the Poisson assumption provides us with a simple, plausible way to transform each BSELL matrix into an HML matrix, and thus to aggregate across stocks.

λ_s used to determine the HIGH, MEDIUM, and LOW cutoffs are specific to each sample. For example, when analyzing a sample of the first 15 minutes of each day, λ_b and λ_s are the mean numbers of buyer-initiated and seller-initiated trades in the first 15 minutes of the trading day.

For each stock, the mapping rules enable us to transform the *n*-by-*n* BSELL matrix into a 3-by-3 (high, medium, low, or HML) matrix. We report three numbers for each cell of the HML matrix: the observed and unexpected percent of intervals belonging to the cell, and the percent of Q_P contributed by the cell. To obtain these numbers, we aggregate over the relevant cells of the BSELL matrix as determined by the mapping rule. Specifically, let o_{ij} be the observed percent of half-hours, u_{ij} be the unexpected percent of half-hours, and Q_{ij} be the Pearson chi-square in cell (i, j) of the BSELL matrix, where

$$o_{ij} = \frac{n_{ij}}{n} \tag{2}$$

$$u_{ij} = n_{ij} - \varepsilon_{ij} \tag{3}$$

$$Q_{ij} = \frac{\left(n_{ij} - \varepsilon_{ij}\right)^2}{\varepsilon_{ij}}, \tag{4}$$

To obtain the percent of observed and unexpected half-hours with, say, LOW BUY and LOW SELL arrivals for a stock, we sum o_{ij} and u_{ij}, respectively, over all cells (i,j) of the BSELL matrix that are mapped into the LOW,LOW cell of the HML matrix. Similarly, to obtain the percent of Q_P contributed by the LOW, LOW cell, we sum Q_{ij} over all cells (i,j) of the BSELL matrix that are mapped into the LOW, LOW cell of the HML matrix, and express this sum as a percent of Q_P. Appendix A provides an illustration of the methodology.

We conduct tests of hypotheses regarding the difference in cell means across different HML tables (e.g. we compare the mean for a particular cell between the table for all days and the table for days with news). To obtain the standard errors of the cell means, we assume that the cell counts follow a Poisson distribution, and estimate a Poisson regression of cell counts on cell and table dummies. Further details on the calculation of standard errors are in Appendix A.

B. Individual Stocks Results

In this section, we present evidence that, at the individual firm level, buyer-initiated and seller-initiated trades are correlated. To test the null hypothesis that the arrivals of buyer-initiated and seller-initiated trades in intervals are statistically independent, we compute, for each stock, the Pearson chi-square statistic.[19] The results are shown in Table 3 for the NYSE stocks (Panel A) and Nasdaq stocks (Panel B). The ticker symbol for each stock is given in column 1, the chi-square statistic is given in column 2, the degrees of freedom is shown in column 3, and the probability value for the chi-square statistic is shown in column 4. The summary statistics at the bottom of each panel show that the null hypothesis of independence is rejected at the 1% level of confidence for 39 or more of the 41 firms in both our NYSE and Nasdaq samples.

Column 5 in the table gives the rank correlation coefficient between buyer-initiated and seller-initiated trades (i.e. the rows and columns of the BSELL matrix) for each stock, and column 6 is the *P*-value for the null hypothesis that the correlation is zero. For all 82 stocks, the null hypothesis of zero correlation is rejected at a high level of significance. For all stocks, the correlation is positive, ranging from 0.25 to 0.77 and averaging 0.49 (for all trades) and 0.57 (for large trades) in the NYSE sample, and ranging from 0.25 to 0.94 and averaging 0.60 (for all trades) and 0.69 (for large trades) in the Nasdaq sample.

The positive association between buyer and seller-initiated trades indicates that markets are two-sided. The typical pattern of two-sidedness is illustrated in Figure 1 for three NYSE stocks and three Nasdaq stocks, those with highest, intermediate and lowest trading frequencies, respectively. The height of each figure shows the percent share of each cell in the HML matrix in the overall chi-square, with "1", "2" and "3" indicating the LOW, MEDIUM and HIGH row or column of the HML matrix, respectively. For the most active NYSE stock, General Motors, and the most active Nasdaq stock, Qualcomm, the HH cell has the dominant share of the overall chi-square. Figure 1 reflects similar patterns for the less active stocks. Recall that the HH cell includes intervals with high numbers of buyer-initiated *and* seller-initiated trades, relative to what would be expected if the trade arrivals followed a Poisson process. Thus, Figure 1 illustrates that the pattern of trade clustering occurs on both the buy and sell sides of the market simultaneously, which indicates that the clustering is generally two-sided.

C. Results for the aggregate of stocks

We next examine whether two-sided trade clustering occurs for stocks assessed collectively. For this purpose, we aggregate the BSELL matrices for individual stocks into one

[19] The Pearson chi-square has previously been used in microstructure studies by, for example, Pasquariello (2001) to examine intra-day patterns in returns and the bid-ask spread in currency markets.

HML matrix, as described in the methodology section. The results are summarized in Table 4 for the NYSE stocks (Panel A) and NASDAQ stocks (Panel B), respectively. We report results for the diagonal cells LL (low buyer-initiated and seller-initiated trade arrivals), MM (medium buyer-initiated and seller-initiated trade arrivals) and HH (high buyer-initiated and seller-initiated trade arrivals). Half-hours with two-sided trade arrivals are represented in one of these three cells. Half-hours with relatively extreme one-sided markets are represented in the HL (high buyer-initiated and low seller-initiated trade arrivals) and LH (low buyer-initiated and high seller-initiated trade arrivals) cells. Half-hours that are neither clearly one-sided nor two-sided are in the (ML, LM) and the (HM, MH) cells (these are not shown in the table).

Each cell contains three statistics, each of which is an average across the stocks: the observed percentage of half-hours in that cell (in the first row), the unexpected percentage of half-hours for that cell (in bold, in the second row), and the percentage contribution of the cell chi-square to the overall chi-square (in the third row). Each panel of Table 4 is divided into four major rows: all half-hours; the first 15 minutes; the last 15 minutes, and the first 15 minutes on news days. Results for tests of hypotheses are presented in the last two tables of each panel.

The results for the different panels of Table 4 show a strikingly consistent pattern of two-sided clustering, even on days with news. Throughout, the incidence of half-hours on the diagonal cells is greater than expected for a random arrival process. Further, the LL and HH cells contribute the dominant share of the overall Pearson chi-square. For example, in panel A, for the sample of all trades, the HH cell has 16.42% of the observed number of half-hours, of which 7.11% are unexpected, and the contribution of this cell to the overall chi-square is 74.61%. The corresponding numbers for the LL cell are 25.51, 7.47 and 8.65, respectively.

In contrast, the number of half-hours with one-sided markets (HL and LH) is less than expected for a Poisson arrival process. For example, in the HL cell, the three entries are 6.19, -6.69 and 2.83 for the observed number of half-hours, the unexpected number, and the contribution to the overall chi-square statistic, respectively. More generally, we find that unexpected buy and sell arrivals are generally negative for all the off-diagonal cells (these results are not reported but are available from the authors). These findings demonstrate an unusually large incidence of intervals with both high buyer-initiated and high seller-initiated trade arrivals, and an unusually small incidence of periods with high trade arrivals on just one side of the market (either on the buy-side or the sell-side). In other words, trade clustering occurs on both sides of the market simultaneously.

The pattern of two-sided clustering also holds for large trades. This may be attributable to the strategic timing of trades by institutional investors executing large orders. By extension, institutional trading in smaller sizes and retail day traders and/or momentum players may explain the pattern in "all trades." This is consistent with the findings of Campbell, Ramadorai and Vuolteenaho (2004) that institutional trading in small sizes is common. Note that slicing and dicing of large institutional orders results in smaller trades, and can dampen the trade clustering to the extent that the sliced and diced orders span into different trading intervals.

Two-sided clustering continues to hold for the first and last 15 minutes of the trading day for both the NYSE stocks and the NASDAQ stocks. Both are periods of heavy volume, and pooling them with the other intervals could spuriously suggest the presence of trade clustering. However, the results show that the pattern of two-sided trade clustering holds even in these heavy volume periods. To illustrate, consider the results for the first 15 minutes for NYSE stocks. *Two-sidedness* is indicated by the positive unexpected percent of intervals for all diagonal cells and negative unexpected percent of intervals in the cells that represent one-sided markets. *Two-sided clustering* is indicated by the large share of the HH cell in the overall chi-square, almost 37%. Thus, two-sided clustering does not appear to be an artifact of pooling relatively low and high volume periods together.

While two-sided markets seem to be the norm in a qualitative sense, the degree of two-sidedness may differ by time of day. In the bottom two tables of each panel, we report results from tests of hypotheses for the mean difference in the observed percent of half-hours between different samples. The first hypothesis relates to the mean difference in the observed percent of half-hours in the diagonal cells (LL, MM, and HH). The second hypothesis relates to the mean difference in the observed percent of half-hours in the off-diagonal cells (LH and HL).

The results of hypotheses tests for different times of the day (penultimate table of each Panel) indicate statistically significant differences in the two-sided pattern. For both NYSE and Nasdaq stocks, the mean percent of intervals in the diagonal cells is generally larger for the whole sample relative to the first and last 15 minutes and, conversely, the mean percent of intervals in the off-diagonal cells (LH and HL) is lower. These results indicate that markets are relatively less two-sided in the first and last 15 minutes, consistent with the idea that opening and closing trades are more likely to be related to news events compared to other trades.

The final major row in Table 4 shows results for the first 15 minutes of days with news. The interesting result is that two-sidedness persists on news days much as it does on all days. For example, considering the sample of large trades, in Panel A (NYSE stocks) the HH cell has the three entries 10.19, 4.73 and 47.81, and in Panel B (NASDAQ stocks), the entries are 16.01, 10.18 and 46.58. In general, the frequency of half-hours in the LH and HL off-diagonal cells is less than expected, and the combined chi-square share of the diagonal cells is about 47% or more, for all trade sizes in both markets. We conclude that, even on news days, the incidence of half-hour windows with high numbers of buyer-initiated and seller-initiated trades is substantially greater than would be expected under a Poisson arrival process.

Turning to the hypotheses tests for news versus non-news days (last table of each Panel), we observe that the difference in the mean percent of half-hours is not statistically significant in most instances in either the NYSE or Nasdaq markets. One exception is for large Nasdaq trades which are *less* one-sided on news days, with 7% lower arrivals in diagonal cells and 2% greater arrivals in the HL, LH cells. Thus, there is little evidence of more one-sided trading sequences following news arrivals.

As we have observed, the evidence of two-sided clustering is similar for the NYSE and Nasdaq markets for all the samples considered: all trades and large trades, various times of the day, and days with news. As such, two-sided trade clustering appears to be a phenomenon that transcends structural differences between these two markets.

D. Order imbalance, total trades, and sidedness

In addition to trade clustering and sidedness, cells in the HML matrix also represent different levels of aggregate trading activity and imbalance in buyer and seller-initiated trades. For example, the number of trades is lower in the LL cells compared to the HH cells; and, controlling for total trades, the imbalance in buyer and seller-initiated trades is greater in the off-diagonal cells than in the diagonal cells. In Easley et al (2005), the absolute imbalance is informative of informed trade arrivals and balanced trades (i.e. total trades minus the absolute imbalance) are informative of uninformed trade arrivals. Is there information content in sidedness (i.e. cells in the HML matrix) beyond order imbalance and total trades?

To address this question, we compare the pattern of sidedness for intervals with more and less imbalance, relative to the median imbalance. Imbalance is defined as the log ratio of absolute imbalance to total trades. Table 5 shows the results. In Panel A of the table, we find that, compared to periods with less imbalance, periods with more imbalance have greater unexpected percent of half-hours and greater chi-square shares in the extreme one-sided cells (i.e. HL and LH) and lower unexpected percent of half-hours and lower chi-square shares in the diagonal cells. The hypotheses test results in Panel B show that, for NYSE (Nasdaq) stocks, there is about 16% (18%) less observations in the diagonal cells and about 10% (12%) more observations in the LH and HL cells in high imbalance periods. These results show that periods with more imbalance are more one-sided and less two-sided, consistent with greater informed and lower uninformed trading arrivals, as in Easley et al (2005). Notably, however, even in periods with high order imbalance, the pattern of two-sided clustering obtains as the unexpected percent of half-hours is negative in the LH and HL cells and positive in the diagonal cells; further, the chi-square share of HH cells exceeds 60%. The implication of these results is that sidedness is not fully captured by the buy-sell imbalance.

Table 5 also shows a comparison of buy-sell arrivals for periods of more and less trades, relative to the median number of trades. The results in Panel B show that periods with more trades are somewhat more two-sided, with about 4% (for NYSE stocks) to 7% (for Nasdaq stocks) greater observations in the diagonal cells. The results are consistent with greater uninformed trading arrivals in periods with more trading. However, even in periods with few trades, the pattern of two-sided markets obtains as the diagonal cells have positive unexpected percent of half-hours along with a combined chi-square share exceeding 50%. Thus, the sidedness variable is informative even after controlling for total trades or the buy-sell imbalance.

E. Summary: Joint arrivals of buyer-initiated and seller-initiated trades

We conclude that buyer and seller-initiated trade arrivals are correlated and cluster together for a wide variety of market scenarios (news and non-news days, different times of the trading day, different trade sizes, and market structures). Since we have also shown that two-sided trade clustering occurs for stocks assessed individually, this finding is not an artifact of pooling firms that trade a lot with firms that trade infrequently. Finally, we find that our measure of sidedness is informative even after accounting for the buy-sell imbalance or total trades.

4. Sidedness, Trade Clustering and Price Volatility

Engle (1996) finds that a shorter inter-trade time interval is associated with higher volatility, implying that prices are more volatile in intervals with increased trade clustering. Consistent with Engle, we show in Appendix B that the clustering of aggregate trades (independent of whether they are buyer or seller-initiated) and volatility are positively associated. In this section, we consider whether market sidedness and trade clustering are associated with price volatility. We first describe the regression methodology used to assess these relationships, and then present our findings.

A. Regression methodology

We use regression analysis to examine the relationship between volatility, sidedness and trade clustering, after controlling for imbalance, number of trades, news, time-of-day and other microstructure effects. Specifically, we regress HILO (the log difference between the maximum and the minimum price in an interval) on five dummy variables that reflect the degree of sidedness and trade clustering (i.e. cells in the 3x3 High-Medium-Low (HML) matrix):

- DUMMY1: equals 1 if the half-hour interval falls in the LL cell
- DUMMY2: equals 1 if the half-hour interval falls in the MM cell
- DUMMY3: equals 1 if the half-hour interval falls in the LH or HL cells
- DUMMY4: equals 1 if the half-hour interval falls in the MH or HM cells
- DUMMY5: equals 1 if the half-hour interval falls in the HH cell

The omitted cells are the LM and ML cells of the HML matrix. DUMMY1, DUMMY2 and DUMMY5 pertain to cells along the diagonal of the HML matrix that represent two-sided markets with increasing levels of activity. In particular, DUMMY5 represents intervals where trades cluster together on both sides of the market. DUMMY3 pertains to the two cells that represent an extreme one sided market, with many trades on one side and few on the other. DUMMY4 pertains to the two cells that represent an intermediate case between extreme two-sidedness (i.e. the HH cells) and extreme one-sidedness (i.e. the HL and LH cells). Referring to Table 1, asymmetric information models predict the highest volatility in the HL and LH cells so that DUMMY3 should have the largest positive coefficient. Models with differential beliefs or information predict volatility to be highest when markets are most two-sided, so that the coefficient of DUMMY5 should have the highest positive coefficient. Finally, portfolio rebalancing implies that coefficients on all five dummy variables are insignificant.

The sidedness dummy variables also incorporate variations in aggregate trading activity and order imbalance. To separate out these effects, we include:

- Log of the number of trades in a half-hour interval
- IMBALANCE: log ratio of the absolute value of order imbalance to the total number of trades. If the imbalance is zero, we add a small number so that the log is defined.

The descriptive statistics in Table 2 show time-of-day effects on volatility, and that volatility is higher on days with news. Accordingly, we include the following dummy variables:

- NEWS: equals 1 on days with news.
- [Open, 15 min after open]: equals 1 for the first 15 minutes of the day.
- [15 to 30 min after open]: equals 1 from 15 to 30 minutes after market open.
- [30 to 15 min before close]: equals 1 from 30 to 15 minutes before market close.
- [15 min before close]: equals 1 for the last 15 minutes of the day.

Higher volatility and higher trading costs are likely to be correlated.[20] Stocks with higher prices may be more liquid and less volatile. Finally, volatility is persistent. Thus, we include:

- Log of the previous day's closing price
- PEBAS: the proportional effective bid-ask half-spread for the interval
- Three lags of HILO.[21]

To further control for stock-specific factors, we also estimate a fixed-effects regression. The reported *T*-statistics are corrected for autocorrelation and heteroskeasticity with the Newey-West procedure, and using 14 lags.

B. Effect of sidedness and trade clustering on volatility

We have computed descriptive statistics for volatility under different conditions of sidedness (i.e. for different cells of the HML matrix). These results (not shown but available

---

[20] See, for example, Subrahmanyam, A., 1994, Circuit breakers and market volatility: A theoretical perspective, Journal of Finance 49,237-254.

[21] For the first-half hour of the day, we use the absolute value of the return from the previous day's closing to the current day's opening price as the first lag of HILO.

from the authors) show that, for both markets, and all trade sizes, the mean and median volatility are generally increasing as we progress from intervals with few trades (the LL cells) to intervals with extreme two-sided trades (the HH cells). Intervals in the HH cells have the highest volatility of all cells, about 60% greater than the volatility in one-sided intervals (the HL and LH cells). For example, for all NYSE trades, the median HILO (times 100) increases from 0.44 in the LL cell to 1.01 in the HH cell. Further, the differences in the mean and median volatilities between the different cells of the HML matrix are statistically significant.

The volatility regression results are given in Table 6, where Panel A is for NYSE stocks, and Panel B is for Nasdaq stocks. In each panel, results for large and all trades are shown separately. Results for the five dummy variables for clustering and sidedness are consistent with the descriptive statistics. For both the NYSE and Nasdaq samples, and for all trade sizes, the dummy coefficient for the LL cell is negative and significant, whereas the coefficients for DUMMY2 (which represents the MM cell) and DUMMY5 (which represents the HH cell) are positive and significant. The DUMMY5 coefficient is the largest in magnitude and the most significant. Thus, all else constant, we observe that volatility increases monotonically as we move diagonally from the LL cells to the HH cells, indicating that volatility is least in two-sided markets with few trades and greatest in two-sided markets with many trades. Further, the DUMMY3 coefficient (representing the LH and HL cells) is positive and significant in three of four cases, but with a magnitude lower than that of DUMMY5. The DUMMY3 coefficient is also smaller then that of DUMMY2 except for large NYSE trades. Thus, volatility is high when markets are extremely one-sided, but not as high as in moderate or extreme two-sided markets. Finally, the DUMMY4 coefficient (representing the MH and HM cells) is always positive and significant, and with magnitudes lower only than that of the DUMMY5 coefficient. These results remain essentially unchanged after we re-estimate the regressions using a dummy variable for each stock except one (the results are not reported but available from the authors).[22]

The above results obtain even after controlling for total trades and order imbalance. We find that volatility is significantly and positively correlated with the number of trades in both exchanges and for all trade sizes. Jones, Kaul and Lipson (1994) and Chan and Fong (2000)

---

[22] When stock-specific fixed effects are introduced, one result different from before is that the DUMMY3 coefficient is either not significant or negative and significant, indicating that volatility is relatively low in extreme one-sided markets.

show a similar result using daily data. The coefficient of IMBALANCE does not have a consistent sign: it is significantly positive (negative) for large (all) NYSE trades, and insignificantly positive (negative) for large (all) Nasdaq trades. Thus, the effect of imbalance on volatility seems difficult to interpret.

Others have found that volatility in the opening and closing minutes of trading is high relative to its value during the rest of the day (see, e.g., Ozenbas, Schwartz, and Wood, 2002). Table 6 shows that volatility is significantly higher in the first 15 minutes of trading, consistent with opening volatility being a price discovery phenomenon, as others have suggested. Holding other variables constant, volatility is significantly lower in the last half-hour of trading, consistent with the descriptive statistics in Table 2.

We find that the coefficient of NEWS is negative and significant. A likely explanation is that the effect of news arrival is largely captured by increased *overnight* price volatility, which is itself accounted for in the regression. Indeed, with lagged values of HILO omitted, the coefficient of NEWS is positive and significant for both Nasdaq and NYSE stocks.

Regarding the remaining variables, HILO is negatively related to the previous day's price (presumably, because stocks with higher prices are generally more liquid and hence less volatile). Trading costs, as represented by PEBAS (the proportional effective half-spread), are positively associated with volatility. Lastly, the three lagged values of HILO are positive and significant, which shows volatility persistence up to 1.5 hours in both markets.

Overall, the relationships described by the regressions depict an economically coherent picture. We find that volatility is highest in periods when many buyer-initiated and seller-initiated trades cluster together, even with order imbalance and total trades accounted for, consistent with trading being motivated by heterogeneous beliefs or information. These results are harder to reconcile with asymmetric information-based models which predict high volatility in one-sided markets. Finally, the results are inconsistent with trading based on portfolio rebalancing since we find a significant association between volatility and the sidedness dummy variables. The adjusted R-squared statistics of around 50% indicate that the independent variables account for an appreciable proportion of the variation in HILO.

5. Trade Clustering and Trading Costs

We now turn to the association between trade clustering and trading costs. Engle and Russell (1994) find evidence of co-movements among duration, volatility, volume, and spread. Consistent with Engle and Russell, we show in Appendix B that greater aggregate trade clustering and trading costs are positively associated. However, market sidedness is likely to be an additional important determinant of trading costs. Therefore, we examine the association between trading costs and the clustering of buyer-initiated and seller-initiated trades. Referring to Table 1, asymmetric information models predict high trading costs when markets are one-sided, whereas the other models have ambiguous predictions on trading costs.

We repeat the regressions described in the previous section with PEBAS (the proportional effective half-spread) as the dependent variable.[23] There are two differences from the previous regressions. We include HILO as an explanatory variable since greater volatility may lead to wider bid-ask spreads by magnifying market maker inventory risks. We also include 3 lags of PEBAS to account for autocorrelation in the bid-ask spread.

We examine descriptive statistics for PEBAS for different cells of the HML matrix. The results (not shown, but available from the authors) show that for both NYSE and Nasdaq stocks, and for large trade sizes, the mean and median trading costs are highest for one-sided markets (the LH, and HL cells). For all trade sizes, PEBAS is generally highest in the HH cells, although close in magnitude to its value in the LH and HL cells. In all cases, the mean and median differences in trading costs between cells are statistically significant.

The trading cost regression results are given in Table 7 for NYSE and Nasdaq stocks, and for large and all trades. For both markets, and for all trade sizes, trading costs are least in two-sided markets. For NYSE stocks, the coefficients for DUMMY1, DUMMY2 and DUMMY5 (that represent the LL, MM and HH cells, respectively) are *negative* and significant in 5 out of 6 possible cases. For Nasdaq stocks, the DUMMY5 coefficient is negative and significant, while the coefficients of DUMM1 and DUMMY2 are negative but not significant. These results indicate that trading costs are relatively low when markets are two-sided, even when there are many trades on both sides (recall that volatility is highest in such cases). In contrast, trading costs are higher in extreme one-sided markets compared to two-sided markets, as indicated by

---

[23] We also have results using PQBAS, the proportional quoted half-spread, as the dependent variable. These results are similar to those using PEBAS and we do not report them (they are available from the authors).

the generally positive and significant coefficient for DUMMY3 that represents the HL and LH cells. The coefficient of DUMMY4 is generally negative, although significant in only one case. Thus, as in the volatility regressions, half-hours represented by DUMMY4 tend to behave more like two-sided than one-sided markets. These results do not change qualitatively even after we include stock specific fixed effects.

The effect of sidedness on trading costs obtains even after controlling for order imbalance and trading activity. The regression results show that trading costs are positively and significantly related to IMBALANCE in all cases. These results are consistent with Corwin and Lipson (2000), who find using daily data that the bid-ask spread increases in response to large order imbalances prior to NYSE trading halts, and Chordia et al (2002) who find that market liquidity is negatively associated with order imbalances at the daily frequency. We also find that trading costs are significantly and negatively related to total trades in all cases.

Turning to the time-of-day dummies, we observe for both Nasdaq and NYSE stocks that trading costs are higher in the first 30 minutes and the last 30 minutes, relative to the rest of the day. While volatility is also higher in the first 15 minutes, relative to the rest of the day, this result obtains even after controlling for volatility. The news day dummy coefficient is positive and significant. Trading costs are negatively related to the prior day's price level, positively related to contemporaneous volatility, and are positively autocorrelated.

Overall, the regression results show that, with order imbalance and total trades, volatility, news, stock-specific fixed effects, time-of-day effects, and other microstructure effects accounted for, trading costs are lower when markets are two-sided compared to one-sided markets. These findings are consistent with the predictions of asymmetric-information based models but are not inconsistent with models based on heterogeneous beliefs or information. The results are inconsistent with trading based on portfolio rebalancing as evidenced by the significant association between trading costs and the sidedness dummy variables. The adjusted R-square is higher than 50% (except for large NYSE trades when it is 26%), indicating that we can explain a large proportion of the variation in PEBAS.

Viewed together, the volatility and trading costs results are most consistent with predictions from models where trading motives are driven by differential beliefs or information. Thus, the results underscore the importance of such motives in stock trading.

6. Additional Investigations

So far, we have shown that, for different exchanges, trade size, time of day, and information conditions, buyer-initiated and seller-initiated trade arrivals are positively correlated, indicating that markets are typically two-sided. In this section, we examine the robustness of our results to a number of potential concerns. Since we identify buyer and seller-initiated trades indirectly, one concern is the accuracy of the Lee-Ready (1991) algorithm for determining the trade direction. In section A, we analyze particular types of trades (e.g. those at the mid-quote) that are more likely to be classified inaccurately, according to prior research. We expect informed trades and one-sided markets to be more prevalent after news releases. Previously, we had assumed that days with the largest unexpected close-to-open returns are news days. In section B, we directly identify news events (earnings releases, for example) and reexamine the evidence for sidedness on these days. In section C, we address the concern that information, and thus one-sided trading sequences, may be short-lived by considering trading intervals shorter than 15 minutes. In section D, we report results from alternative methodologies for determining sidedness, and for alternative measures of volatility. Finally, we discuss whether our results may be due to the presence of stale limit orders.

A. The effect of errors in classifying the trade direction

Ellis, Michael and O'Hara (2000) show, for Nasdaq stocks, and Peterson and Sirri (2003) find, for NYSE stocks, that the Lee-Ready (1991) algorithm is accurate between 81% and 93% of the time. However, the algorithm is less accurate for trades that are inside the quotes and, in particular, for trades at the mid-quote; in addition, accuracy is lower for large stocks and for the post-decimalization period. Accordingly, we repeat our analysis for these types of trades. In our sample, trades that are at the mid-quote (inside quotes but not at the mid-quote) constitute about 10% (27%) of NYSE trades and 8% (36%) of Nasdaq trades. Since decimalization occurred in January 2001, we choose June 2000 as a pre-decimalization period. To analyze large and small stocks, we split the sample into the 20 largest and smallest stocks, based on their market capitalization as of January 3, 2003. The results are reported in Table 8.

Panel A of Table 8 shows that markets are two-sided for all types of trades, with positive (negative) unexpected arrivals and large (small) chi-square contributions in the diagonal (off-diagonal) cells of the HML matrix. The hypothesis tests show that, compared to all trades, when trades are inside quotes and at the mid-quote, there are more observations on *both* the diagonal and off-diagonal cells. The same is also true when comparing pre- and post-decimalization trades. Thus, these results do not indicate a bias towards either more one-sided or more two-sided markets due to trade classification errors, or due to decimalization. The lack of a bias from decimalization is reassuring since trade sizes decreased substantially after decimalization,[24] suggesting the possibility that as large orders were broken up more, markets became more two-sided after decimalization (assuming large trades to be more one-sided). However, the results from Panel A of Table 8 indicate that this is not the case. Indeed, an examination of large trades (results not reported) in the pre-decimalization period further confirms that two-sided markets are typical even prior to decimalization. Finally, hypotheses tests comparing the 20 largest and smallest stocks show that large stock trades are moderately more two-sided than small stock trades, consistent with the intuition of market practitioners.

We next investigate whether differences in the buy-sell arrivals for different trade types are reflected in the way sidedness and clustering are associated with volatility and trading costs. Panel B of Table 8 shows results from regressions of HILO and trading costs on sidedness for trades inside quotes and at the mid-quote; results for other control variables are not reported for brevity. The measure of trading costs is PEBAS for trades at inside quotes and PQBAS, or the proportional quoted half-spread, for trades at the mid-quote since PEBAS is zero for trades at the mid-quote. The results are generally consistent with previous results: HILO is highest in intervals with two-sided clustering (i.e. the HH cells) and trading costs are highest in one-sided intervals (i.e. the LH and HL cells). The only exception is that, for NYSE trades at the mid-quote, PQBAS is highest in the MH, HM and HH cells rather in the LH and HL cells.

Overall, our results for both NYSE and Nasdaq stocks are robust to trade classification errors and to the choice of sample period. In addition, Peterson and Sirri (2003) show that the Lee-Ready (1991) algorithm works best if no lags are incorporated when matching trades to prevailing quotes, a procedure we have followed throughout this paper.

---

[24] For example, Chordia, Sarkar and Subrahmanyam (2005) document that, after decimalization, the average daily number of trades for the largest NYSE stocks increased from about 2,400 to almost 4,000.

B. Clustering and sidedness on days with corporate news events

Proper identification of news events is essential to finding evidence of informed trade arrivals and one-sided markets that are more likely to occur following news events. We searched major publications for news relating to earnings, dividends, mergers and acquisitions, share repurchases or stock splits, or changes in credit ratings. We found 39 NYSE stocks and 28 Nasdaq stocks with news in these categories for our sample period. Panel A of Table 9 shows that, on days with news, stocks have significantly higher close-to-open returns or ACLOP, volatility or HILO, volume or VOL, number of trades or #TR, and effective spreads or PEBAS. However, for Nasdaq stocks, PEBAS is not significantly different on news and non-news days. These results are identical to those found when using the value of ACLOP to identify news days.

Panel B of Table 9 shows the distribution of buyer-initiated and seller-initiated trades for the first 15 minutes of days with news. As previously, we find negative unexpected arrivals in the HL and LH cells, with a chi-square contribution of about 18% for these cells. Further, unexpected arrivals are generally positive in the diagonal cells except for the MM cell for NYSE stocks; the chi-square share of the LL and HH cells add up to about 33% for NYSE stocks and 50% for Nasdaq stocks. Panel C of Table 9 shows that for Nasdaq stocks, there is about 15% more observations on diagonal cells on news days, indicating that markets are *more* two-sided at this time. For NYSE stocks, sidedness is not significantly different for news and non-news days. These results are similar to what we found previously.

Panel D of Table 9 shows results from regressions of volatility and trading costs on a news dummy and sidedness; results for other control variables are not reported for brevity. The news dummy is not significant, whereas in previous regressions it was negative and significant. Most important, results for the sidedness dummies are robust, demonstrating that our results are not sensitive to alternative identifications of news days.

C. Results using 1-minute windows

Thus far, we have examined sidedness for 15 and 30 minute windows and found no evidence of one-sided markets. However, if information is generally short-lived, then one-sided trading sequences may be observed over short trading windows after news arrival. To examine this possibility, we examine windows of 10, 5, 3 and 2 minutes and find that markets remain two-sided even on news days. In Table 10, we examine trade arrivals over 1-minute windows for the first 15 minute of trading days. For all days, Panel A of the table shows that markets are two-sided with negative (positive) unexpected arrivals in off-diagonal (diagonal) cells. However, for news days, we find evidence of one-sided markets for NYSE stocks. In particular, unexpected arrivals are positive for LH and HL cells and negative for LL and MM cells. Note, however, that the magnitude of unexpected arrivals in the LH and HL cells is only about 1%; further, the combined chi-square share of these cells is about 13%. In addition, for HH cells, unexpected arrivals are positive though small (at 0.39%) and the chi-square share exceeds 46%. Finally, there is no evidence of one-sided trading in Nasdaq stocks. Panel B of Table 10 shows that for NYSE (Nasdaq) stocks there are about 19% (5%) greater observations on diagonal cells on all days relative to news days. Thus, there is evidence of one-sided trading on news days using 1-minute windows, although the evidence is weak and mostly confined to NYSE stocks.

Panel C of Table 10 shows regressions of trading cost and volatility on sidedness for the first 15 minutes of the day. The results are similar to those reported for half-hour intervals. In particular, HILO is highest for two-sided intervals with clustering (i.e. HH cells), and not for one-sided intervals (i.e. the LH and HL cells), even for NYSE stocks. In unreported results, we multiplied the sidedness dummies with the NEWS dummy. We found that, for NYSE stocks, volatility is positively associated with one-sided intervals only for news days, and not for non-news days. Thus, one-sided trading leads to enhanced volatility, consistent with asymmetric information-based models. However, volatility is most strongly associated with trade arrivals that are two-sided and clustered.

D. Alternative methodology for estimating sidedness

Our evidence for two-sided markets is based on the HML matrix, where the low, medium and high cutoffs are derived using the square root of the mean number of buyer and seller-initiated trades for each stock. As an alternative, we now use the actual standard deviation of trade arrivals. Specifically, we calculate the z-score for a stock in an interval as the number of buyer or seller-initiated trades minus the sample mean, and divided by the sample standard deviation.[25] We then take the correlation between the z-scores for buyer and seller-initiated trades. A large and positive correlation indicates two-sided markets. We find that, for all

---

[25] We thank Eugene Kandel for suggesting this approach to us. The results are similar if we subtract the z-score for the "market," defined as the average z-score for all stocks in the interval.

intervals, the average correlation is 0.49 for NYSE stocks and 0.60 for Nasdaq stocks. For the first 15 minutes of a day, the correlation drops to 0.35 for NYSE stocks and 0.51 for Nasdaq stocks. For the first 15 minutes of news days, the correlation is 0.32 for NYSE stocks and 0.54 for Nasdaq stock. These results are consistent with two-sided clustering, with relatively more one-sided trading during the first 15 minutes of days. As previously, news arrivals do not make a substantial difference to the sidedness.

E. Alternative definitions of volatility

We define volatility as the log ratio of the maximum to the minimum price in a half-hour interval, which may include the bid-ask bounce. To address this concern, we repeat our analysis after defining volatility as the log ratio of the maximum to the minimum mid-quote. The results (not shown but available from the authors) are similar to those reported earlier. In particular, volatility is highest in the HH cells for both NYSE and Nasdaq stocks. We also define volatility as the sum of 1-minute squared returns in a half-hour interval and obtain qualitatively similar results. We conclude that our results are robust to alternative definitions of volatility.

F. Effect of stale limit orders

The arrival of good news may prompt an influx of market buy orders that hit standing limit orders before they can be withdrawn, causing a clustering of buyer-initiated trades.[26] Similarly, bad news may cause a clustering of seller-initiated trades. Note, however, that our results show trade clustering on *both* sides of the market. Thus, for stale limit orders to cause two-sided markets, good news must typically follow bad news, or vice versa, within our measurement interval. However, we find that two-sided markets obtain even with a short measurement window of 1-minute, which makes it unlikely that our results are primarily due to the presence of stale limit orders.

7. Conclusion

We have examined the pattern of trade arrivals in the two major U.S. market centers, the NYSE and Nasdaq. Of primary interest is the tendency for trades to cluster together in relatively brief intervals (from 30 minutes to 1 minute) within a trading day. We have observed a greater prevalence of both high and low volume intervals (and a paucity of intermediate volume intervals) relative to what would be expected if trade execution were a random arrival process. The clustering is clear for both the NYSE and Nasdaq markets.

We have further assessed the extent to which the clusters are one-sided (buyers only or sellers only) or two-sided (buyers and sellers are present together). The evidence points to the latter. For the array of market conditions that we have considered (marketplace, trade size, time of day, and information environment), buyer-initiated and seller-initiated trade arrivals are positively correlated, which means that trade bursts are two-sided. Particularly striking is the extent to which two-sidedness continues to prevail on days with news release. Apparently, markets are efficient in the sense that prices move rapidly into new trading ranges within which some participants are looking to buy shares, and others are seeking to sell shares.

Could the two-sided clustering be an artifact of pooling large and small trades? No, we find the effect for large trades separately. Is it explained by pooling high and low volume intervals? No, we find a similar pattern in the opening and closing minutes of the trading day, times when volume typically spikes up on a daily basis. Is the appearance of two-sided trade clusters explained by changes in the informational environment? No, we observe two-sided clustering on both news and non-news days. Could the findings be attributable to our having aggregated large and small volume stocks? Once again, the answer is negative – the aggregation procedure that we have used normalizes for trading volume and the application of the tests to individual stocks further shows that two-sided trade bursts are prevalent for each stock.

After having established the prevalence of two-sided trade clustering, we considered the association of these bursts with price volatility and trading costs. With order imbalance, number of trades, news arrival, time-of-day effects and share price controlled for, we observe that trade bursts are associated with higher volatility, and that the trading costs are generally higher when markets are one-sided than when they are two-sided.

Our sidedness variable contains information not fully captured by order imbalance and total trades. We find that markets are two-sided even in periods of relatively high imbalance; further, our sidedness variables are highly significant in explaining volatility and trading costs after controlling for imbalance and total trades. We suggest that sidedness is informative because it depends on the relation between the distributions of buyer-initiated and seller-initiated trades, whereas order imbalance is a summary measure of these distributions.

---

[26] We thank Joel Hasbrouck for suggesting this possibility that stale limit orders may cause clustering.

Earlier Dufour and Engle (2000), among others, found evidence that aggregate trades cluster in time. The evidence that buyer-initiated and seller-initiated trade arrivals are positively correlated under a wide variety of market conditions suggests that clustering is not simply attributable to asymmetric information. We find evidence of one-sided trading sequences in the first 15-minute of news days when we examine windows of 1-minute duration. This suggests that the effect of news arrivals is short-lived, and that trading based on asymmetric information is a relatively small share of aggregate trading activity. This inference is consistent with the relatively small values of the probability of informed trading estimated by Easley et al (2005).

We suggest that two-sided trade clustering may arise from participants having different information or divergent beliefs, and in price and quantity discovery therefore being complex, dynamic processes. Imperfect quantity discovery, which is associated with a two-sided latent demand to trade, suggests that trade bursts are not necessarily attributable to the arrival of new fundamental information per se. Rather, participants on both sides of the market may simultaneously wish to trade but do not reveal their orders until some event (e.g., the arrival of enough other orders and trades) animates them to do so. This interpretation of two-sided inter-temporal trade clustering leads to several implications for trader behavior and market structure.

First, the prevalence of two-sidedness for all trade sizes underscores the importance of divergent beliefs or information as motives for trading. Previous empirical research has focused on divergent expectations as a motive for trading treasury securities (Fleming and Remolona, 1999). Our findings underscore the possibility that the same motive exists for equity trading.

Second, the evidence that trading occurs in bursts suggests that some orders at least are portable in time and that, at any point in time, a two-sided, *latent* demand to trade exists. To the extent that this is indeed the case, trade clusters may not be attributable only to the release of fundamental news that generates a fresh demand to trade. Rather, something can occur in the marketplace that leads participants, on both sides of the market, to step forward, take existing orders out of their pockets, and trade. Whatever seeds or animates the process, trading appears to gain strength as the latent demands of both buyers and sellers are turned into active orders. As orders are activated, the time duration between trades decreases and a trade burst occurs.

Third, finding that markets are commonly two-sided indicates that natural buyers and natural sellers (the investors) are generally present in the market at the same time and, consequently, that they should, in principle, be able to supply liquidity to each other. This might suggest that intermediaries are not strictly needed in the marketplace (at least for larger issues). Nevertheless, intermediaries are widely recognized to have an important function to perform with regard to matters such as price discovery and the provision of immediacy. Our analysis implies that they may have another important role – to animate trading.[27]

Fourth, the existence of a sizable, two-sided, latent demand to trade implies that market structure is not enabling buyers and sellers to meet each other appropriately, and that an important source of liquidity (latent liquidity) is not being adequately exploited. Thus, further examination of trade clusters and the magnitude of latent liquidity would be desirable.

We leave for future research a more complete analysis of what might spark bouts of intensified trading, but nevertheless suggest the following. Accurate price and quantity discovery are difficult to achieve in an environment where participants have different share valuations, and where their desires to trade are not fully revealed to the market.[28] Orders that are "held in traders' pockets" represent a latent demand to trade. Whatever flushes latent demand into the market can cause the two-sided trading bursts that we observe.[29]

[27] The term "animation" applies specifically to an intermediary contacting potential buyers and sellers and/or by actually triggering trades themselves. More generally, exchange floor traders and market makers are widely recognized as being market facilitators to the extent that they actively bring buyers and sellers together. In part, they might do so by stimulating book building and by triggering trades.

[28] Traders bring their orders carefully to the market for two reasons. First, they are concerned about market impact costs. Second, being subjected to trader performance evaluations, they are reluctant to trade at "undesirable" prices (i.e., buy above or sell below the volume weighted average price for a trading session). Both considerations lead participants to bring their orders to the market in smaller pieces that are executed over an extended period of time.

[29] Circumstantial evidence suggests that the latent demand to trade may be consequential. In the current environment, it is well known that large institutional traders typically slice and dice their big orders for execution over extended periods of time (up to a day or more). This reality is reflected in the fact that average trade size has declined form a peak of 2303 shares in 1988 to 393 shares in 2004. Concurrently, block-trading volume (trades of 10,000 shares or more) on the NYSE has dropped sharply from 51.1% of reported volume in 1988 to 31.9% in 2004 (www.nysedata.com/factbook). New trading facilities such as Liquidnet and Pipeline have been designed with explicit reference to this deficiency. Moreover, evidence suggests that (1) institutional participants believe it takes months, not days or less, for price dislocations to be repaired in the market, (2) their demand for immediacy is largely attributable to the dynamics of the marketplace (which include, e.g., front running), and (3) portfolio managers commonly give their traders a day or more to work their orders (see Schwartz and Steil, 2002) for discussion and further references).

References

1. Admati, A.R. and P. Pfleiderer, 1988, A Theory of Intraday Patterns: Volume and Price Variability, Review of Financial Studies, 1, 3-40.

2. Anderson, E. W., E. Ghysels and J. L. Juergens, 2004, Do Heterogeneous Beliefs Matter for Asset Pricing? Working Paper, University of North Carolina at Chapel Hill.

3. Antunovich, P. and A. Sarkar, forthcoming, Fifteen Minutes of Fame? The Market Impact of Internet Stock Picks, Journal of Business.

4. Bamber, L. D., O. E. Barron and T. L. Stober, 1999, Differential Interpretations and Trading Volume, Journal of Financial and Quantitative Analysis, 3, 369-386.
5. Benston, G., and R. Hagerman, 1974, Determinants of Bid-Asked Spreads in the Over-The-Counter Market, Journal of Financial Economics, 1, 353-364.

6. Campbell, J. Y., Ramadorai, T. and T. Vuolteenaho, 2004, Caught on Tape: Predicting Institutional Ownership with Order Flow, Working Paper, Harvard University.

7. Chan, K. and W. Fong, 2000, Trade Size, Order Imbalance, and the Volatility-Volume Relation, Journal of Financial Economics, 57, 247-273.

8. Chordia, T., R. Roll and A. Subrahmanyam, 2002, Order Imbalance, Liquidity and Market Returns, Journal of Financial Economics, 65, 111-130.

9. Chordia, T., A. Sarkar and A. Subrahmanyam, 2005, The Joint Dynamics of Liquidity, Returns and Volatility, Working Paper, Anderson School of Management, UCLA.

10. Corwin, S. and M. L. Lipson, 2000, Order Flow and Liquidity Around NYSE Trading Halts, Journal of Finance, 55, 1771-1801.

11. Dufour, A. and R. F. Engle, 2000, Time and the Price Impact of a Trade, Journal of Finance, 55, 2467-2498.

12. Easley, D., N. M. Kiefer and M. O’Hara, 1996, Cream Skimming or Profit Sharing, Journal of Finance, 51, 811-833.

13. Easley, D., N. M. Kiefer and M. O’Hara, 1997a, The Information Content of the Trading Process, Journal of Empirical Finance, 4, 159-186.

14. Easley, D., N. M. Kiefer and M. O’Hara, 1997b, One Day in the Life of a Very Common Stock, Review of Financial Studies, 10, 805-835.

15. Easley, D., R. F. Engle, M. O’Hara and L. Wu, 2005, Time-Varying Arrival Rates of Informed and Uninformed Trades, Working Paper.

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

17. Engle, R. F., 1996, The Econometrics of Ultra-High Frequency Data, Discussion Paper 96-15, Department of Economics, University of California, San Diego.

18. Engle, R. F. and J. R. Russell, 1994, Forecasting Transaction Rates: The Autoregresive Conditional Duration Model, NBER Working paper No. 4966.

19. Fleming, M. and E. M. Remolona, 1999, Price Formation and Liquidity in the U.S. Treasury Market: The Response to Public Information, Journal of Finance, 54, 1901-1915.

20. Foster, F. D. and S. Viswanathan, 1994, Strategic Trading with Asymmetrically Informed Traders and Long-Lived Information, Journal of Financial and Quantitative Analysis, 29, 4, 499-518.

21. Frankel, J. F. and K. Froot, 1990, Chartists, Fundamentalists and Trading in the Foreign Exchange Market, American Economic Review, 80, 181-185.

22. Grundy, B. D. and M. McNichols, 1990, Trade and the Revelation of Information Through Prices and Direct Disclosure, Review of Financial Studies, 2, 495-526.

23. Hall, A. D. and N. Hautsch, 2004, A Continuous Time Measurement of the Buy-Sell Pressure in a Limit Order Book Market, Working Paper, University of Copenhagen.

24. Harris, M. and A. Raviv, 1993, Differences of Opinion Make a Horse Race, Review of Financial Studies, 6, 473-506.

25. Hasbrouck, J., 1991, Measuring the Information Content of Stock Trades, Journal of Finance, 46, 179-207.

26. He, Hua and Jiang Wang, 1995, Differential Information and Dynamic Behavior of Stock Trading Volume; Review of Financial Studies; 8, 919-972.

27. Hendershott, T. and C. M. Jones, 2005, Island Goes Dark: Transparency, Fragmentation, and Regulation, Review of Financial Studies, 18, 743-793.

28. Jones, Charles M., G. Kaul, and M. L. Lipson, 1994, Transactions, Volume and Volatility; Review of Financial Studies, 7, 631-651.

29. Kandel, E. and N. D. Pearson, 1995, Differential Interpretation of Public Signals and Trade in Speculative Markets, Journal of Political Economy, 103, 831-872.

30. Kim, O. and R. E. Verrecchia, 1994, Liquidity and Volume Around Earnings Announcements, Journal of Accounting and Economics, 17, 41-67.

31. Kyle, A.S., 1985, Continuous Auctions and Insider Trading, Econometrica, 53, 1315-1335.

32. Lee, C. and M. Ready, 1991, Inferring Trade Directions from Intraday Data, Journal of Finance, 46, 733-746.

33. Llorente, G., R. Michaely, G. Saar and J. Wang, 2002, Dynamic Volume-Return Relation of Individual Stocks, Review of Financial Studies, 4, 1005-1047.

34. Ozenbas, D, R. Schwartz and R. Wood, 2002, Volatility in U.S. and European Equity Markets: An Assessment of Market Quality, International Finance, 5, 437-461.

35. Pasquariello, P., 2001, The Microstructure of Currency Markets: An Empirical Model of Intra-day Return and Bid-Ask Spread Behavior; Working Paper, University of Michigan.

36. Peterson. M. and E. Sirri, 2003, Evaluation of the Biases in Execution Cost Estimation Using Trade and Quote Data, Journal of Financial Markets, 6, 259-280.

37. Schwartz, R. and B. Steil, 2002, Controlling Institutional Trading Costs: We Have Met the Enemy, and it is Us, The Journal of Portfolio Management, 28, 39-49.

38. Shalen, C. T., 1993, Volume, Volatility and the Dispersion of Beliefs, Review of Financial Studies, 6, 405-434.

39. Subrahmanyam, A., 1994, Circuit breakers and market volatility: A theoretical perspective, Journal of Finance, 49, 237-254.

40. Wang, J., 1993, A Model of Intertemporal Asset Prices Under Asymmetric Information, Review of Financial Studies, 60, 249-282.

41. Wang, J., 1994, A Model of Competitive Stock Market Trading, Journal of Political Economy, 102, 127-168.

Appendix A

Methodology for Hypotheses Tests

We provide an illustration of the methodology for estimating the joint distribution of buyer-initiated and seller-initiated trades, and then discuss details of the Poisson regression used for hypothesis testing in Table 4.

A. Illustration of methodology used to estimate joint distribution of buyer-initiated and seller-initiated trade arrivals

To illustrate, suppose we have a stock that averages 4 large buy trades and 3 large sell trades per half-hour. Further, suppose that there is a maximum of 8 large buys and 6 large sell trades in the half-hour intervals in our sample. We first construct a 9-by-7 buy-sell matrix (that includes additional cells for no large buyer-initiated and for no large seller-initiated trades). Each cell, ranging from (0,0) to (8,6), gives the number of half-hours with the specific buy-sell combination for that cell. The Pearson chi-square Q_P is then computed as described in (1) of the text. The 9-by-7 matrix is mapped into a 3-by-3 matrix as follows:

- Half-hours with 2 large buy trades (the mean of 4 minus the standard deviation of 2) or less are mapped into a LOW BUY cell; and half-hours with 1 (the mean of 3 minus the square root of 3 rounded down) or zero large sell trades are mapped into a LOW SELL cell.
- Half-hours with 6 or more large buy trades are mapped into a HIGH BUY cell; and half-hours with 5 or more large sell trades are mapped into a HIGH SELL cell.
- Half-hours with more than 2 but less than 6 large buy trades are mapped into a MEDIUM BUY cell; and half-hours with more than 1 but less than 5 large sell trades are mapped into a MEDIUM SELL cell.

The arrival frequency in the LOW, LOW cell is obtained by summing o_{ij} over i=0,1,2 and j=0,1. The contribution of the LOW, LOW cell to Q_P is obtained by summing Q_{ij} over i=0,1,2 and j=0,1, and expressing the sum as a percent of Q_P. Numbers for the other cells are obtained in similar fashion.

B. Tests of hypotheses

We conduct tests of hypotheses regarding the difference in cell means across different HML tables (e.g. comparing the mean for cell HH between the table for NYSE stocks and the table for Nasdaq stocks). To obtain the standard errors of the cell means, we assume that the cell counts for different stocks and tables follow a Poisson distribution.

Denote n_{tijs} as the cell count for row i and column j in table t for stock s. We compare two tables at a time, so t=1,2. Let n_{ts} denote the sum of cell counts for table t and stock s. Further, let I_{ijt} denote an indicator variable that equals 1 for row i, column j and table t, and is zero otherwise. Then, we estimate a Poisson regression model as follows:

$$\log(n_{tijs}) = \beta_0 + \beta_1 \log(n_{ts}) + \sum_{t=1}^{2}\sum_{i=1}^{3}\sum_{j=1}^{3} \beta_{tji} I_{tji} + u_{tji} \qquad \text{(A1)}$$

β_1 is assumed to be 1, so that $\log(n_{st})$ is interpreted as a so-called offset variable; it normalizes the fitted cell means to a percent of the total cell count for the stock and the table. u_{tji} is the error term.

The regression (A1) is estimated by maximizing the log-likelihood function L= $\sum_{i=1}^{738} l_i$ with respect to the regression parameters. l_i is the log-likelihood for the i-th observation, and the total number of observations is 738 (equal to the number of stocks (41) times the number of cells (9) times the number of tables (2)). For the Poisson distribution, l_i has the form:

$$l_i = n_i \log(\mu_i) - \mu_i \qquad \text{(A2)}$$

L is maximized using a ridge-stabilized Newton-Raphson algorithm (details available from the authors). In all cases, the algorithm converged.

We estimate regression (A1) to obtain the cell means and standard errors for each table, which are then used to calculate t-statistics for testing hypotheses about mean differences in the usual way. For example, the estimated *percent* means for cell HH in tables 1 and 2 are given by:

$$\hat{\mu}_{133} = \hat{\beta}_0 + \hat{\beta}_{133} A_{133}$$

$$\hat{\mu}_{233} = \hat{\beta}_0 + \hat{\beta}_{233} A_{233} \qquad \text{(A3)}$$

Suppose that the corresponding standard errors are estimated as $\hat{s}e_{133}$ and $\hat{s}e_{233}$. Then, to test whether the mean for cell HH in table 1 is different from that in table 2, the t-statistic is:

$$t = \frac{\hat{\mu}_{133} - \hat{\mu}_{233}}{\sqrt{\left(\hat{s}e_{133}\right)^2 + \left(\hat{s}e_{233}\right)^2}} \tag{A4}$$

The degree of freedom for the t-statistic is 736, since the total number of observations is 738 (equal to the number of stocks (41) times the number of cells (9) times the number of tables (2)).

We also compare the sum of cell means (e.g. the mean for cell LH plus the mean of cell HL) across tables. In this case, to obtain the standard errors, we assume for simplicity that the variance of the sum of means is equal to the sum of the variances of the means.

Appendix B
Results on Aggregate Trade Clustering

In this appendix, we suppress the distinction between buyer-initiated and seller-initiated trades, and investigate clustering for trades in the aggregate. Trade clustering is defined as an unusually high number of trade arrivals in particular half-hour intervals. We first describe the empirical methodology used to determine aggregate trade clustering and then present our findings. We discuss results for all trades as well as for large trades, for the first and last 15 minutes of the trading day, for the first 15 minutes of days with news, and for NYSE and Nasdaq stocks. Finally, we discuss results of regressions of price volatility and trading costs on trade clustering. We do not report results, but they are available from the authors.

A. Methodology

Our objective is to assess the observed and unexpected percent of half-hours with high, medium and low trade arrivals. To this end, we first count, for each stock, the number of half-hour windows for which a particular number of trades (e.g. two trades) was observed, and record these in a trade (TRADE) row vector. Let n_j be the observed number of half-hours with exactly j trades (e.g. j=2 trades), and let ε_j be the expected number of half-hours in cell j of the TRADE vector. $\varepsilon_j=P_jn$, where P_j is the probability of observing n_j, and $n=\sum_j n_j$ is the total number of half-hours in the sample. P_j is obtained under the assumption that trades follow a Poisson arrival process, with parameter λ equal to the mean number of trades per half-hour for a stock for a particular sample.[1] The observed and expected percentage of half-hours with exactly j trades are $o_j=n_j/n$ and $e_j=\varepsilon_j/n$, respectively, so that the unexpected percentage is $u_j=o_j-e_j$. Finally, for cell j, we define $Q_j=(n_j-\varepsilon_j)^2/\varepsilon_j$ and, summing over all cells of the TRADE vector, we define the statistic $Q_P=\sum_j Q_j$. Under the null hypothesis that the cell proportions are given by P_j, Q_P has an asymptotic chi-square distribution with (C-1) degrees of freedom, where C is the number of columns in the TRADE vector. Then, the chi-square contribution (in percent) of cell j to Q_P is $\chi_j=100*(Q_j/Q_P)$.

---

[1] Our results also hold under alternative assumptions about P_j (e.g., the probability of observing n_j is the same for all j). Later, when we examine a matrix of buyer-initiated and seller-initiated trades, rather than a TRADE vector, the probability is obtained directly from the row and column sums, and no assumption is necessary as to whether the trade arrivals follow a particular distribution (such as Poisson).

To draw conclusions about a trade burst pattern for stocks in aggregate, we transform the 1-by-*n* TRADE vectors for the individual stocks into 1-by-3 High, Medium, and Low (HML) vectors that can be aggregated across stocks. Because trading activity (and, hence, the size of the TRADE vector) differs across stocks, we standardize each stock-specific TRADE vector so that stocks with different arrival rates are comparable. For each stock, a half-hour interval with f trades is mapped into the:

- LOW cell if f <= Rounddown(λ-$\sqrt{\lambda}$)
- HIGH cell if f > Roundup (λ+$\sqrt{\lambda}$)
- MEDIUM cell in all other cases.

To obtain the observed and unexpected percent of trade arrivals in the 1-by-3 HML vector, and the contribution of each cell to Q_P, we aggregate o_j, e_j and χ_j over the relevant cells of the TRADE vector as determined by the mapping rule. We then average each of the three cells in the HML vector across the stocks in our sample. For instance, to obtain o_j for the LOW trade arrivals for a specific stock, we sum o_j over all cells of the TRADE vector for that stock that are mapped into the LOW cell of the HML vector. After repeating this process for each stock, we compute the mean of o_j across stocks.

C. Results on aggregate trade clustering

We first discuss results on trade clustering for NYSE stocks. We find that the LOW (henceforth L) and HIGH (henceforth H) cells have the largest observed and unexpected percent of half-hours, and that the H cells have the largest contribution to the overall chi-square. This is true for all trades as well as for the sample of large trades. For example, considering the large trades, 45.09 percent and 28.09 percent, of the observed number and the unexpected number, respectively, of half-hours occur in the L cell, and 23.18 percent and 8.18 percent, of the observed and unexpected number, respectively, of half-hours occur in the H cell. The H cell accounts for practically all of the table chi-square. In contrast, the sign of the unexpected percent of half-hours in the MEDIUM (henceforth M) cells is *negative*. Consequently, our classification of half-hours according to the frequency of trades within them is U-shaped (being relatively low in the M cells, and relatively high in the L and H cells). The dominant contribution of the H cells to the overall chi-square shows that half-hour intervals with high trade arrivals are more prevalent than would be expected if trade arrival was a random process. We call this phenomenon "clustering" or "trade bursts."

Results for the first and the last 15-minute samples show that the U-shaped pattern persists in the heavy volume periods themselves. For example, considering all trades during the first 15 minutes of a trading day, the percent of observed and unexpected 15-minute intervals is, respectively, 43.14 and 27.06 in the L cell, and 32.68 and 16.52 in the H cell. As before, the H cell accounts for the major share (66.51 percent) of the overall chi-square. Thus, the evidence for trade bursts does not rely on pooling relatively low and high volume periods.

The evidence further shows that trade bursts occur in the first 15-minute period on news days much as they do on non-news days. Thus, for large trades, the percent of the observed and unexpected number of 15-minute intervals is, respectively, 44.15 and 27.93 in the L cell, and 24.29 and 8.19 in the H cell, and the H cell accounts for 69.05 percent of the overall chi-square. Thus, the concentration of trading volume does not appear to be an artifact of pooling informationally intense periods with informationally sparse periods.

Finally, the evidence of trade bursts for the NYSE and Nasdaq markets are strikingly similar. As such, trade clustering appears to transcend structural differences between these two markets.

In conclusion, it is clear that the half-hour windows pattern continues to be U-shaped, with spikes in the number of half-hours with very few trades (Low) and with many trades (High). It is also clear that the large number of half-hours in the High cell is the dominant contributor to the overall chi-square; this constitutes, in our terminology, a trade burst.

D. Regression results

We use regression analysis to examine the relationships between volatility and trade clustering, after controlling for microstructure effects, news, and time-of-day effects. Specifically, we regress HILO on dummy variables that reflect the degree of trade clustering. The dummy variables refer to cells in the 1x3 High-Medium-Low (HML) trade vector:

- DUMMY1: equals 1 if the half-hour interval falls in the Low (L) cells
- DUMMY2: equals 1 if the half-hour interval falls in the High (H) cells

The omitted cell is the Medium (M) cell of the HML vector. A negative coefficient for DUMMY1 indicates that a low level of trade arrivals is associated with low volatility, and a positive coefficient for DUMMY2 indicates that volatility is higher when a trade burst occurs.

We first consider the association of price volatility and aggregate trade clustering. Descriptive statistics show that, for both markets and for all trade sizes, the mean and median volatility increase as we progress from half-hour intervals with low trade arrivals to intervals with high trade arrivals. For instance, the mean high-low range for all trades in NYSE stocks is 0.57 for half-hours in the L cells; it increases to 0.73 in the M cells and to 1.06 in the H cells.[2] We find that the mean and median volatility in the M and H cells are significantly different from those in the L cells. The volatility regression results are consistent with the descriptive statistics. We find, for both the NYSE and Nasdaq samples, that the dummy coefficient for the L cell is negative and significant, and that for the H cell it is positive and significant. Volatility is clearly highest in the half-hours where a trade burst has occurred.

Next, we discuss the association of trading costs and aggregate trade clustering. Descriptive statistics for PEBAS under different conditions of clustering show that, for NYSE stocks, and for the sample of all trades and for the sample of large trades, the mean and median trading costs increase as we move from the L, to the M, to the H cells. For instance, the mean PEBAS for NYSE stocks is 0.0821, 0.0891 and 0.0977, while the median PEBAS is 0.0574, 0.0639 and 0.0701, in the L, M and H cells, respectively. We find that these differences are statistically significant. The results for the Nasdaq stocks are qualitatively similar to those for the NYSE stocks. PEBAS is lowest in the L cells but, unlike the NYSE stocks, trading costs are highest in the M cells, and they decrease as we move from the M cells to the H cells.

The trading cost regression results are broadly consistent with the previously discussed descriptive statistics. We find for both the NYSE and Nasdaq samples that trading costs are lowest in the L cells, as indicated by the significantly negative dummy coefficient for the L cells. For NYSE stocks, trading costs are highest in the H cells, whereas for Nasdaq stocks, trading costs are highest in the M cells, which is somewhat surprising. Perhaps, this result reflects the increased liquidity supplied by market makers or public limit order traders in the Nasdaq market.

We conclude that trading costs are higher in intervals where many trades cluster, and lower in intervals with few trades.

---

[2] Similarly, the median high-low range increases monotonically from 0.47 (for L) to 0.87 (for H).

Table 1: Models, Predictions, and Findings

Model	Prediction	Consistent With Our Findings?
1. Some investors have superior information about asset value (Wang , 1993, 1994; Llorente, Michaely, Saar and Wang, 2002)	• One-sided markets • Trade clustering on one side of the market • Higher volatility when markets are one-sided • Higher trading cost when markets are one-sided	• No • No • No • Yes
2. Investors have different private information signals (Grundy and McNichols, 1989; Shalen, 1993; He and Wang, 1995) and/or different interpretations of public news (Kim and Verrecchia, 1994; Kandel and Pearson, 1995)	• Two-sided markets • Trade clustering on both sides of the market • Higher volatility when markets are two-sided • Ambiguous effect on trading costs	• Yes • Yes • Yes • Yes
3. Investors trade to rebalance their portfolios (Wang , 1994; He and Wang, 1995; Llorente, Michaely, Saar and Wang, 2002)	• Two-sided markets • No implication for clustering • No relation between sidedness and volatility • No relation between sidedness and trading costs	• Yes • No • No • No

Table 2: Descriptive Statistics

The table shows cross-sectional means for 41NYSE and 41Nasdaq stocks during January 2 to May 31 2003. The NYSE and Nasdaq stocks are matched according to their closing price and market value on December 31 2002. MCAP (in $million) is the market capitalization and PRICE is the closing price on January 2 2003. ACLOP is the absolute value of the excess return from the previous day's closing price to the current day's opening price. HILO is log of the highest to the lowest price in an interval. Intervals are of 30-minutes duration, with the first and last half-hours broken up into two 15-minute intervals. The remaining measures are computed separately for all trade sizes (*All*) and large trades (*Large*), defined as those in the top 10 percentile of the dollar value of trades of a stock. VOLUME and TRADES are total volume and number of trades per interval. BUYS and SELLS are the number of buy-triggered and sell-triggered trades per interval, determined using the Lee-Ready (1991) algorithm; numbers for the 15 minute intervals are multiplied by two for consistency. PQBAS (PEBAS) is the average proportional quoted (effective) bid-ask half-spread in an interval. PQBAS is the quoted bid-ask spread divided by 2M, where M is the quote mid-point. PEBAS is Q*(P- M)/M, where P is the trade price, and Q is +1 (-1) for a buyer (seller) initiated trade. Estimates for HILO, RETURN, PQBAS, and PEBAS are multiplied by 100. News days for a stock are the 30 percentile of days with the largest value of ACLOP. ** indicates whether the means are significantly different, at the one percent level or less, between news and non-news days, or between the two open and close 15-minute intervals and the middle half-hours.

Panel A: NYSE stocks

	All days	News days	Non-news days	Open to 15 min after open	15 to 30 min after open	Middle half-hours	30 to 15 min before close	15 min before to close
OBS	54,226	17,958	36,074	4,167	4,167	45,883	4,171	4,172
MCAP	4,742	4,742	4,742	4,742	4,742	4,742	4,742	4,742
PRICE	21.5571	21.5571	21.5571	21.5571	21.5571	21.5571	21.5571	21.5571
ACLOP	0.7693	1.6737**	0.3074	0.7693	0.7693	0.7693	0.7693	0.7693
HILO	0.7667	0.8424**	0.7286	1.0529**	0.8367**	0.7046	0.4915**	0.5718**
VOLUME	123,404	134,002**	118,071	183,317**	200,279**	109,826	159,515**	249,605**
TRADES	90	94**	88	101**	123**	85	111**	141**
All buys	46.7466	49.5646**	45.3546	53.8612**	64.8672**	43.8294	57.6794**	74.7573**
All sells	38.9485	41.0951**	37.8882	43.8181**	52.9273**	36.7423	47.8020**	59.7845**
PQBAS, All	0.1902	0.2042**	0.1832	0.3477**	0.2472**	0.1827	0.1589**	0.1779
PEBAS, All	0.0888	0.0932**	0.0866	0.1654**	0.1099**	0.0854	0.0768**	0.0841
Large buys	4.8790	5.3708**	4.6361	7.2470**	7.8646**	4.1980	7.0211**	12.3656**
Large sells	3.6793	4.0742**	3.4843	5.5169**	5.9588**	3.1963	5.2105**	8.6588**
PQBAS, Large	0.2388	0.2530**	0.2315	0.3869**	0.2887**	0.2308	0.1992**	0.2148
PEBAS, Large	0.1182	0.1267**	0.1139	0.2738**	0.1333**	0.1113	0.0915**	0.0984**

Table 2: Descriptive Statistics

Panel B: Nasdaq stocks

	All days	News days	Non-news days	Open to 15 min after open	15 to 30 min after open	Middle half-hours	30 to 15 min before close	15 min before to close
OBS	54,415	18,190	36,225	4,176	4,185	46,042	4,184	4,187
MCAP	4,441	4,441	4,441	4,441	4,441	4,441	4,441	4,441
PRICE	21.3517	21.3517	21.3517	21.3517	21.3517	21.3517	21.3517	21.3517
ACLOP	0.9563	2.0087**	0.4199	0.9563	0.9563	0.9563	0.9563	0.9563
HILO	0.8991	0.9953**	0.8508	1.5093**	1.0021**	0.8020	0.5864**	0.7986
VOLUME	275,629	312,486**	257,122	589,980**	466,347**	242,256	293,117**	488,898**
TRADES	200	221**	189	355**	313**	179	233**	342**
All buys	99.5823	111.2761**	93.7878	176.3642**	157.4753**	89.3830	115.9299**	172.9444**
All sells	93.7916	105.0772**	88.1994	168.8250**	146.4660**	84.2443	110.1714**	159.7457**
PQBAS, All	0.0924	0.0958**	0.0906	0.1370**	0.1071**	0.0895	0.0865	0.0972**
PEBAS, All	0.0646	0.0671**	0.0634	0.0991**	0.0745**	0.0625	0.0599**	0.0678**
Large buys	10.3083	12.3430**	9.3000	23.2394**	17.7187**	8.7572	11.7675**	22.6318**
Large sells	9.1419	10.9079**	8.2669	21.2314**	15.8802**	7.7606	10.4708**	19.3749**
PQBAS, Large	0.0941	0.0968**	0.0927	0.1313**	0.1032**	0.0910	0.0863**	0.1011**
PEBAS, Large	0.0705	0.0725**	0.0695	0.0996**	0.0768**	0.0685	0.0638**	0.0724**

Table 3: Tests of Independence of Buy and Sell Trade Arrivals

The table shows, for each stock, results for tests of the null hypothesis that buyer- and seller-initiated trades in a half-hour interval are statistically independent. The statistical measures are computed separately for the sample of All trade sizes (*All*) and the sample of large trades (*Large*), defined as those in the top 10 percentile of the dollar value of trades of a stock. Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. We count the number of half-hour windows for which each combination of buy and sell triggered trades (e.g. two buy trades and one sell trade) was observed, and record them in a buy-sell (BSELL) matrix. Our null hypothesis is that the rows and columns of the BSELL matrix are independent. To test the hypothesis, we use the Pearson chi-square statistic Q_P, which is equal to $Q_P = \sum_i \sum_j \frac{(n_{ij} - \varepsilon_{ij})^2}{\varepsilon_{ij}}$, where n_{ij} is the observed frequency of buy-triggered and sell-triggered trade arrivals in row *i* and column *j*, the expected frequency (under the null hypothesis of independence) is $\varepsilon_{ij} = \frac{n_{i.} n_{j.}}{n}$, $n_{i.} = \sum_j n_{ij}$ is the sum for row *i*, $n_{.j} = \sum_i n_{ij}$ is the sum for column *j*, and $n = \sum_i \sum_j n_{ij}$ is the overall total. When the rows and columns are independent, Q_P has an asymptotic chi-square distribution with (R-1)(C-1) degrees of freedom (DOF), where R is the number of rows and C is the number of columns in the matrix. The table also shows the correlation (*Corrln*) between the rows and columns of the BSELL matrix. Panel A shows results for NYSE stocks and Panel B for Nasdaq stocks. The sample is 41 NYSE and 41 Nasdaq stocks during January 2 to May 31 2003. The NYSE and Nasdaq stocks are matched according to their closing price and market value on December 31 2002.

Table 3: Tests of Independence of Buy and Sell Trade Arrivals

Panel A: NYSE stocks

	All trade sizes					Large trades				
Ticker	Chi-square	DOF	P	Corrln	P	Chi-square	DOF	P	Corrln	P
ABX	23,065.90	1,050	0.0000	0.62	0.0000	3,250.87	1,050	0.0002	0.54	0.0000
ALA	19,038.11	667	0.0000	0.60	0.0000	8,764.46	667	0.0000	0.67	0.0000
AMD	55,071.49	2,520	0.0000	0.65	0.0000	12,326.77	2,520	0.0000	0.68	0.0000
APC	34,208.16	1,672	0.0000	0.62	0.0000	6,690.34	1,672	0.0000	0.64	0.0000
AWE	25,017.57	2,150	0.0078	0.36	0.0000	8,428.28	2,150	0.0000	0.56	0.0000
AXE	13,450.94	306	0.0000	0.50	0.0000	5,119.95	306	0.0000	0.60	0.0000
CC	23,659.07	1,332	0.0000	0.54	0.0000	4,915.95	1,332	0.0000	0.61	0.0000
CD	21,659.24	1,457	0.0403	0.27	0.0000	3,438.00	1,457	0.0001	0.36	0.0000
CI	40,002.09	1,976	0.0000	0.62	0.0000	16,286.22	1,976	0.0000	0.74	0.0000
CMH	6,869.41	357	0.0003	0.37	0.0000	4,942.58	357	0.0000	0.54	0.0000
CMX	14,445.61	750	0.0000	0.47	0.0000	8,568.16	750	0.0000	0.56	0.0000
CUZ	3,421.07	121	0.0008	0.27	0.0000	928.04	121	0.0064	0.27	0.0000
CVC	23,205.91	1,026	0.0000	0.56	0.0000	5,450.26	1,026	0.0000	0.60	0.0000
DO	22,502.08	825	0.0000	0.69	0.0000	3,692.99	825	0.0000	0.56	0.0000
EW	14,026.89	384	0.0000	0.35	0.0000	3,617.67	384	0.0000	0.51	0.0000
FNF	20,828.40	864	0.0000	0.56	0.0000	9,597.78	864	0.0000	0.68	0.0000
FVB	10,072.29	783	0.0000	0.51	0.0000	12,253.04	783	0.0000	0.75	0.0000
GGP	9,443.35	306	0.0000	0.37	0.0000	5,743.47	306	0.0000	0.54	0.0000
GM	49,987.13	5,265	0.0000	0.47	0.0000	16,757.62	5,265	0.0000	0.65	0.0000
GMH	27,217.50	676	0.0000	0.47	0.0000	8,547.14	676	0.0000	0.62	0.0000
HHS	7,634.66	256	0.0000	0.45	0.0000	2,246.97	256	0.0000	0.49	0.0000
HRC	25,153.77	624	0.0000	0.77	0.0000	7,655.32	624	0.0000	0.68	0.0000
HU	3,836.69	130	0.0001	0.26	0.0000	1,704.07	130	0.0000	0.31	0.0000
IGL	6,792.27	368	0.0029	0.37	0.0000	3,764.51	368	0.0000	0.58	0.0000
IRF	19,662.36	1,170	0.0000	0.58	0.0000	8,538.34	1,170	0.0000	0.65	0.0000
KEM	7,112.00	294	0.0002	0.29	0.0000	1,400.06	294	0.0054	0.41	0.0000
KSE	16,664.31	1,088	0.0000	0.55	0.0000	8,737.21	1,088	0.0000	0.65	0.0000
LSI	22,911.39	1,221	0.0000	0.49	0.0000	8,448.16	1,221	0.0000	0.61	0.0000
NUE	17,337.12	1,394	0.0000	0.59	0.0000	11,627.18	1,394	0.0000	0.67	0.0000
OGE	12,240.08	357	0.0000	0.44	0.0000	2,989.53	357	0.0000	0.50	0.0000
PDG	24,981.77	784	0.0000	0.64	0.0000	3,769.00	784	0.0000	0.61	0.0000
RAD	21,935.50	756	0.0000	0.53	0.0000	3,308.39	756	0.0000	0.45	0.0000
RDC	27,190.02	1,271	0.0000	0.67	0.0000	6,308.42	1,271	0.0000	0.64	0.0000
RGA	2,752.55	156	0.0529	0.29	0.0000	606.28	156	0.0112	0.32	0.0000
SLB	50,285.35	2,964	0.0000	0.67	0.0000	9,591.56	2,964	0.0000	0.68	0.0000
SVM	8,179.91	315	0.0000	0.37	0.0000	1,875.07	315	0.0006	0.43	0.0000
TCB	17,744.59	598	0.0000	0.51	0.0000	5,239.03	598	0.0000	0.58	0.0000
TTN	23,068.52	806	0.0000	0.61	0.0000	11,981.23	806	0.0000	0.76	0.0000
UNM	61,658.50	2,160	0.0000	0.75	0.0000	11,632.20	2,160	0.0000	0.74	0.0000
URI	6,872.54	234	0.0000	0.36	0.0000	3,262.78	234	0.0000	0.48	0.0000
WDR	7,050.31	304	0.0000	0.25	0.0000	1,951.51	304	0.0000	0.46	0.0000

	No. of stocks	No. sig at 1%	Sum of chi-square	Sum of DOF	Avg. corrln
All trade sizes	41	39	848,256.40	41,737	0.49
Large trades	41	40	265,956.44	41,737	0.57

Table 3 (continued)

Panel B: Nasdaq stocks

	All trade sizes					Large trades				
Ticker	Chi-square	DOF	P	Corrln	P	Chi-square	DOF	P	Corrln	P
ATML	97,091.79	9,785	0.0000	0.71	0.0000	24,503.47	9,785	0.0000	0.82	0.0000
BEAS	168,872.00	23,393	0.0000	0.74	0.0000	31,896.84	23,393	0.0000	0.83	0.0000
CBCF	6,805.33	272	0.0000	0.34	0.0000	2,520.57	272	0.0001	0.46	0.0000
CIEN	155,810.00	19,912	0.0000	0.66	0.0000	33,002.85	19,912	0.0000	0.84	0.0000
CMCSK	143,427.00	19,845	0.0000	0.67	0.0000	22,387.60	19,845	0.0000	0.75	0.0000
COGN	56,298.45	2,346	0.0000	0.75	0.0000	11,342.36	2,346	0.0000	0.74	0.0000
COMS	56,027.96	4,176	0.0000	0.67	0.0000	15,495.81	4,176	0.0000	0.75	0.0000
EXPD	40,502.48	2,704	0.0000	0.43	0.0000	7,188.25	2,704	0.0000	0.60	0.0000
FAST	39,968.23	1,886	0.0000	0.53	0.0000	8,869.98	1,886	0.0000	0.66	0.0000
GSPN	57,125.60	2,964	0.0000	0.69	0.0000	11,809.93	2,964	0.0000	0.71	0.0000
HBAN	29,019.91	1,974	0.0000	0.44	0.0000	5,056.41	1,974	0.0000	0.51	0.0000
HCBK	7,631.42	340	0.0003	0.33	0.0000	2,189.61	340	0.0003	0.37	0.0000
ICBC	9,003.43	552	0.0003	0.30	0.0000	3,029.18	552	0.0000	0.38	0.0000
ICST	58,725.60	3,685	0.0000	0.70	0.0000	10,773.19	3,685	0.0000	0.70	0.0000
IMCL	132,974.00	7,912	0.0000	0.93	0.0000	38,372.34	7,912	0.0000	0.94	0.0000
INTU	168,014.00	17,550	0.0000	0.78	0.0000	45,735.92	17,550	0.0000	0.92	0.0000
IPCR	12,382.51	285	0.0000	0.40	0.0000	1,586.16	285	0.0018	0.35	0.0000
JNPR	177,884.00	25,456	0.0000	0.80	0.0000	40,148.41	25,456	0.0000	0.89	0.0000
LRCX	57,229.65	3,520	0.0000	0.66	0.0000	11,568.46	3,520	0.0000	0.68	0.0000
MOLX	58,139.97	3,410	0.0000	0.58	0.0000	12,432.02	3,410	0.0000	0.75	0.0000
NBTY	36,996.20	1,400	0.0000	0.64	0.0000	8,441.78	1,400	0.0000	0.65	0.0000
NTAP	144,715.00	15,561	0.0000	0.72	0.0000	23,523.02	15,561	0.0000	0.84	0.0000
NXTL	197,095.00	40,232	0.0000	0.66	0.0000	42,008.54	40,232	0.0000	0.84	0.0000
PDCO	35,772.32	2,150	0.0000	0.57	0.0000	13,872.37	2,150	0.0000	0.73	0.0000
PHCC	44,817.69	1,720	0.0000	0.66	0.0000	15,048.02	1,720	0.0000	0.75	0.0000
QCOM	246,312.00	55,144	0.0000	0.77	0.0000	55,043.84	55,144	0.0000	0.89	0.0000
QTRN	27,488.12	1,680	0.0000	0.52	0.0000	12,707.45	1,680	0.0000	0.68	0.0000
RFMD	150,938.00	15,730	0.0000	0.77	0.0000	41,500.97	15,730	0.0000	0.89	0.0000
ROST	65,871.47	4,216	0.0000	0.67	0.0000	15,763.41	4,216	0.0000	0.74	0.0000
RSLN	25,261.20	1,023	0.0000	0.39	0.0000	6,037.12	1,023	0.0000	0.46	0.0000
SAFC	30,358.62	2,250	0.0000	0.51	0.0000	7,380.90	2,250	0.0000	0.56	0.0000
SPLS	94,439.75	10,791	0.0000	0.48	0.0000	15,427.49	10,791	0.0000	0.70	0.0000
SSCC	41,609.50	3,540	0.0000	0.45	0.0000	9,809.43	3,540	0.0000	0.60	0.0000
SUNW	209,115.00	39,104	0.0000	0.56	0.0000	37,103.64	39,104	0.0000	0.82	0.0000
SWFT	27,371.19	1,404	0.0000	0.52	0.0000	6,698.26	1,404	0.0000	0.56	0.0000
SYMC	183,706.00	19,305	0.0000	0.84	0.0000	42,684.77	19,305	0.0000	0.93	0.0000
TECD	50,391.19	2,491	0.0000	0.70	0.0000	19,537.38	2,491	0.0000	0.86	0.0000
TRST	4,558.38	196	0.0013	0.33	0.0000	332.50	196	0.0549	0.25	0.0000
USON	11,017.46	483	0.0000	0.35	0.0000	2,453.97	483	0.0001	0.33	0.0000
WFMI	63,767.78	3,240	0.0000	0.70	0.0000	18,648.94	3,240	0.0000	0.85	0.0000
YHOO	223,032.00	40,068	0.0000	0.79	0.0000	50,976.21	40,068	0.0000	0.89	0.0000

	No. of stocks	**No. sig at 1%**	**Sum of chi-square**	**Sum of DOF**	**Avg. corrln**
All trade sizes	41	41	3,447,567.18	413,695	0.60
Large trades	41	40	784,909.35	413,695	0.69

Table 4: Distribution of buyer and seller-initiated trades

Each cell of the table reports, averaged over stocks, and for a particular buy-and-sell-trade arrival combination, the observed and **unexpected (in bold)** percent of half-hours, and the chi-square statistic of the cell as a percent of the overall chi-square. Numbers are reported for the following buy-and-sell-trade arrival combination: low buyer-initiated and low seller-initiated trade arrivals (LL), medium buyer-initiated and medium seller-initiated trade arrivals (MM), high buyer-initiated and low seller-initiated trade arrivals (HL), low buyer-initiated and high seller-initiated trade arrivals (LH), and high buyer-initiated and high seller-initiated trade arrivals (HH). Statistics are shown for different times of the day, and on days with news. The statistical measures are computed separately for the sample of All trade sizes (*All*) and the sample of large trades (*Large*), defined as those in the top 10 percentile of the dollar value of trades of a stock. Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. News days for a stock are the 30 percentile of days with the largest values of ACLOP, the absolute value of the log excess returns from the previous day's closing price to the current day's opening price.

Details of the calculations are as follows. We first count the number of half-hour windows for which each combination of buy and sell triggered trades (e.g. two buy trades and one sell trade) was observed, and record them in a buy-sell (BSELL) matrix. Let n_{ij} denote the observed number of half-hours in cell (i, j) of the BSELL matrix. The expected number of half-hours in cell (i, j) is $\varepsilon_{ij} = \frac{n_{i.}n_{.j}}{n}$, where $n_{i.} = \sum_j n_{ij}$ is the sum for row i, $n_{.j} = \sum_i n_{ij}$ is the sum for column j, and $n = \sum_i \sum_j n_{ij}$ is the overall total. The Pearson chi-square in cell (i, j) is $Q_{ij} = \frac{(n_{ij} - \varepsilon_{ij})^2}{\varepsilon_{ij}}$ and the overall table chi-square Q_P is given by $Q_P = \sum_i \sum_j \frac{(n_{ij} - \varepsilon_{ij})^2}{\varepsilon_{ij}}$, so that the chi-square contribution of cell (i, j) to Q_P is $\chi_{ij} = \frac{Q_{ij}}{Q_P}$. Finally, let o_{ij}, e_{ij} and u_{ij} be the observed, expected and unexpected percent of half-hours in cell (i, j), where $o_{ij} = \frac{n_{ij}}{n}$, $e_{ij} = \frac{\varepsilon_{ij}}{n}$ and $u_{ij} = o_{ij} - e_{ij}$.

For each stock, the BSELL matrix is then mapped into a 3-by-3, High-Medium-Low (HML) matrix as follows. Assume that buy trades follow a Poisson arrival process, with parameter λ_b equal to the mean of the number of buy trades for the stock for a particular sample (e.g., all days or days with news). Then, for each stock and each sample, a half-hour interval with n_b buy trades is mapped into the:

- LOW BUY cell if n_b <= Rounddown(λ_b-$\sqrt{\lambda_b}$)
- HIGH BUY cell if n_b > Roundup (λ_b+$\sqrt{\lambda_b}$)
- MEDIUM BUY cell in all other cases.

An identical procedure is carried out for sell trades, under the assumption that sell trades follow a Poisson arrival process, with parameter λ_s equal to the sample mean of the number of sell trades for the stock.

To obtain the observed and unexpected percent of trade arrivals, and the contribution of each cell to Q_P., we aggregate o_{ij}, e_{ij} and χ_{ij} over the relevant cells of the BSELL matrix as determined by the mapping rule. For example, to obtain these numbers for the HH cell, we sum o_{ij}, u_{ij} and χ_{ij} over all cells (i,j) of the BSELL matrix that are mapped into the HH cell of the HML matrix.

Results from hypotheses tests are shown under the heading, *Mean Differences in Observed Percent of Half-Hours*. We show t-statistics and p-values for the null hypotheses that the difference in mean (μ) of observed percent of half-hours between different times of the day, and between news and non-news days, is zero for (1) the diagonal cells and (2) the HL and LH cells. The comparison is for stocks commonly traded in the two samples. The standard errors used to compute the t-statistics are obtained from a Poisson regression of cell counts on cell and table dummies, as described in Appendix A of the text. ** (*) indicates whether the means are significantly different, at the one (five) percent level or less. The sample is 41 NYSE (Panel A) and 41 Nasdaq stocks (Panel B) during January 2 to May 31 2003. The NYSE and Nasdaq stocks are matched according to their closing price and market value on December 31 2002.

Table 4 (continued)

Panel A. NYSE stocks

Distribution of buyer-initiated and seller-initiated trades

	All trade sizes					Large trades				
	H,L	**L,H**	**H,H**	**L,L**	**M,M**	**H,L**	**L,H**	**H,H**	**L,L**	**M,M**
All half-hours										
Obs %	6.19	6.16	16.42	25.51	8.84	2.85	2.80	10.96	24.20	17.25
Unexp %	**-6.69**	**-6.80**	**7.11**	**7.47**	**1.09**	**-5.81**	**-5.80**	**6.38**	**7.52**	**2.29**
Chi-sq share %	2.83	3.24	74.61	8.65	1.02	1.29	1.28	92.25	1.38	0.44
First 15 minutes, all days										
Obs %	8.45	8.71	13.85	21.55	8.34	6.96	6.84	9.76	20.44	11.75
Unexp %	**-4.35**	**-3.98**	**4.18**	**4.44**	**0.30**	**-2.91**	**-3.37**	**4.07**	**2.08**	**-0.14**
Chi-sq share %	8.73	9.57	36.83	16.42	2.98	3.65	3.45	69.47	3.90	2.71
Last 15 minutes, all days										
Obs %	10.58	11.13	13.83	20.15	6.82	6.94	6.65	11.22	22.54	11.46
Unexp %	**-3.13**	**-2.54**	**2.95**	**2.82**	**0.10**	**-4.32**	**-3.99**	**4.25**	**5.23**	**1.18**
Chi-sq share %	11.32	13.72	29.40	14.91	3.09	4.86	6.66	59.02	6.60	2.52
First 15 minutes, news days										
Obs %	9.33	7.89	12.48	21.52	8.11	6.31	6.24	10.19	21.95	10.98
Unexp %	**-3.76**	**-3.95**	**3.28**	**4.29**	**-0.14**	**-4.14**	**-3.51**	**4.73**	**2.38**	**-0.54**
Chi-sq share %	8.56	9.20	22.47	18.59	5.71	5.69	7.46	47.81	6.06	4.47

Mean difference in observed percent of half-hours, for different times of day

	All trade sizes				Large trades			
Null hypothesis	**No. stocks commonly traded**	**Estimate**	**T-statistic**	**P-value**	**No. stocks commonly traded**	**Estimate**	**T-statistic**	**P-value**
All half-hours – First 15 minutes								
μ(LL+MM+HH)=0 μ(LH+HL)=0	41	7.04** -4.82**	6.66 -7.35	0.0000 0.0000	41	10.18** -8.16**	9.67 -14.04	0.0000 0.0000
All half-hours – Last 15 minutes								
μ(LL+MM+HH)=0 μ(LH+HL)=0	41	9.98** -9.36**	9.71 -12.81	0.0000 0.0000	41	7.18** -7.95**	6.64 -13.67	0.0000 0.0000

Mean difference in observed percent of half-hours, for first 15 minutes of all days and news days

	All trade sizes				Large trades			
Null hypothesis	**No. stocks commonly traded**	**Estimate**	**T-statistic**	**P-value**	**No. stocks commonly traded**	**Estimate**	**T-statistic**	**P-value**
First 15 minutes, all days –First 15 minutes, news days								
μ(LL+MM+HH)=0 μ(LH+HL)=0	41	1.69 -0.04	0.84 -0.03	0.4034 0.9730	41	-1.19 1.24	-0.58 1.12	0.5617 0.2650

Table 4 (continued)

Panel B. Nasdaq stocks

Distribution of buyer-initiated and seller-initiated trades

	All trade sizes					Large trades				
	H,L	**L,H**	**H,H**	**L,L**	**M,M**	**H,L**	**L,H**	**H,H**	**L,L**	**M,M**
All half-hours										
Obs %	7.51	7.51	19.99	36.75	3.44	3.02	2.79	13.33	35.07	11.33
Unexp %	**-9.16**	**-9.45**	**9.35**	**9.92**	**0.66**	**-8.08**	**-8.10**	**8.21**	**10.63**	**2.65**
Chi-sq share %	3.35	3.10	75.20	9.58	0.61	0.88	0.92	93.63	1.52	0.35
First 15 minutes, all days										
Obs %	9.80	9.73	18.83	33.39	3.01	4.24	4.55	14.35	42.41	6.46
Unexp %	**-7.29**	**-7.08**	**7.18**	**7.66**	**0.48**	**-8.82**	**-8.55**	**8.68**	**10.48**	**1.78**
Chi-sq share %	10.49	9.07	33.50	23.91	2.03	3.40	2.69	69.33	8.81	2.25
Last 15 minutes, all days										
Obs %	12.79	13.78	18.02	28.18	2.18	6.01	5.75	14.35	32.82	7.20
Unexp %	**-4.51**	**-4.64**	**4.13**	**4.99**	**-0.03**	**-7.30**	**-7.33**	**6.88**	**8.97**	**1.22**
Chi-sq share %	12.63	14.22	27.23	22.71	1.78	4.86	4.85	58.15	9.45	2.32
First 15 minutes, news days										
Obs %	9.77	9.24	20.02	37.19	1.37	3.67	3.00	16.01	50.30	4.48
Unexp %	**-8.74**	**-8.05**	**7.99**	**8.46**	**-0.33**	**-10.51**	**-10.60**	**10.18**	**12.63**	**1.71**
Chi-sq share %	10.47	9.88	24.46	32.46	1.39	4.60	3.58	46.58	19.63	4.63

Mean difference in observed percent of half-hours, for different times of day

	All trade sizes				Large trades			
Null hypothesis	**No. stocks commonly traded**	**Estimate**	**T-statistic**	**P-value**	**No. stocks commonly traded**	**Estimate**	**T-statistic**	**P-value**
All half-hours – First 15 minutes								
μ(LL+MM+HH)=0	41	4.96**	4.16	0.0000	41	-3.83**	-3.02	0.0026
μ(LH+HL)=0		-4.51**	-6.45	0.0000		-2.98**	-6.33	0.0000
All half-hours – Last 15 minutes								
μ(LL+MM+HH)=0	41	11.81**	10.53	0.0000	41	5.36**	4.55	0.0000
μ(LH+HL)=0		-11.54**	-14.25	0.0000		-5.96**	-11.05	0.0000

Mean difference in observed percent of half-hours, for first 15 minutes of all days and news days

	All trade sizes				Large trades			
Null hypothesis	**No. stocks commonly traded**	**Estimate**	**T-statistic**	**P-value**	**No. stocks commonly traded**	**Estimate**	**T-statistic**	**P-value**
First 15 minutes, all days –First 15 minutes, news days								
μ(LL+MM+HH)=0	41	-3.21	-1.36	0.1739	41	-7.21**	-2.79	0.0054
μ(LH+HL)=0		0.41	0.30	0.7661		2.12*	2.54	0.0112

Table 5: Distribution of buyer-initiated and seller-initiated trades, for periods of more and less imbalance, or more and less trades

Panel A of the table reports for each cell, averaged over stocks, the observed and **unexpected (in bold)** percent of half-hours, and the chi-square statistic of the cell as a percent of the overall chi-square. Numbers are reported for the HL, LH, HH, LL, and MM cells of the HML matrix, where the first (second) letter refers to buyer (seller) initiated trade arrivals, and L, M, H refer to Low, Medium, and High, respectively. Statistics are shown for periods with more and less imbalance, or with more and less trades, relative to the median imbalance or number of trades. Imbalance is the log ratio of the absolute imbalance to total trades. Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. Results from hypotheses tests are shown in Panel B. We show *t*-statistics and *p*-values for the null hypothesis of zero mean difference (μ=0) in the observed percent of intervals in (1) the diagonal cells and (2) the HL and LH cells. The comparison is for stocks commonly traded in the two samples. The standard errors used to compute the *t*-statistics are from a Poisson regression of cell counts on cell and table dummies. ** (*) indicates whether the means are significantly different, at the one (five) percent level or less. The sample is 41 NYSE and 41 Nasdaq stocks during January 2 to May 31 2003. The NYSE and Nasdaq stocks are matched using their closing price and market value on December 31 2002.

Panel A: Distribution of buyer-initiated and seller-initiated trades

	All trade sizes, NYSE stocks					All trade sizes, Nasdaq stocks				
	H,L	L,H	H,H	L,L	M,M	H,L	L,H	H,H	L,L	M,M
Periods with more imbalance										
Obs %	7.40	10.71	14.07	23.33	6.77	10.96	11.76	15.89	32.07	1.40
Unexp %	**-6.28**	**-2.57**	**4.35**	**4.52**	**0.02**	**-5.93**	**-5.34**	**5.13**	**4.95**	**-1.19**
Chi-sq share %	4.08	8.35	60.76	9.37	1.41	6.46	7.11	61.27	11.11	0.35
Periods with less imbalance										
Obs %	2.70	2.39	19.89	30.51	13.47	3.53	3.22	23.87	41.36	6.56
Unexp %	**-9.51**	**-10.05**	**11.16**	**12.94**	**4.55**	**-12.76**	**-13.07**	**13.56**	**15.32**	**3.06**
Chi-sq share %	1.38	1.56	73.77	14.19	1.80	1.58	1.49	73.31	14.95	1.37
Periods with more trades										
Obs %	6.43	6.70	15.65	23.65	10.77	8.71	7.88	18.07	32.26	5.30
Unexp %	**-5.90**	**-6.00**	**6.67**	**5.99**	**0.76**	**-7.22**	**-7.86**	**7.86**	**7.31**	**0.10**
Chi-sq share %	4.52	5.12	65.54	8.37	1.93	6.54	5.57	60.95	10.31	1.02
Periods with less trades										
Obs %	6.94	6.95	12.26	20.56	14.86	10.82	11.67	14.12	29.25	8.66
Unexp %	**-3.93**	**-4.08**	**4.32**	**4.98**	**1.29**	**-4.35**	**-3.94**	**3.66**	**4.83**	**0.21**
Chi-sq share %	6.67	6.46	57.65	8.39	6.13	11.96	12.13	20.26	24.47	8.05

Panel B. Mean difference in observed percent of half-hours

	All trade sizes, NYSE stocks				All trade sizes, Nasdaq stocks			
Null hypothesis	No. stocks	Estimate	T-stat	P-value	No. stocks	Estimate	T-stat	P-value
Periods with more imbalance– Periods with less imbalance								
μ(LL+MM+HH)=0	41	-16.19**	-36.61	0.0000	41	-18.38**	-39.46	0.0000
μ(LH+HL)=0		9.89**	49.24	0.0000		12.28**	52.15	0.0000
Periods with more trades– Periods with less trades								
μ(LL+MM+HH)=0	40	3.91**	9.20	0.0000	41	6.84**	14.90	0.0000
μ(LH+HL)=0		-0.26	-1.19	0.2357		-0.35	-1.46	0.1445

Table 6: Regressions of Volatility on Sidedness and Clustering

The table shows results from a regression of the volatility on dummy variables for sidedness and clustering. Statistics are reported separately for the sample of All trade sizes (*All*) and the sample of large trades (*Large*), defined as those in the top 10 percentile of the dollar value of trades of a stock in the sample. The sample is 41 NYSE (Panel A) and 41 Nasdaq (Panel B) stocks during January 2 to May 31 2003. The NYSE and Nasdaq stocks are matched according to their closing price and market value on December 31 2002. The proxy for volatility, the dependent variable, is HILO, equal to the log of the highest to the lowest price in a half-hour interval. HILO is regressed on dummy variables for sidedness and clustering. They refer to cells in the 3x3 High-Medium-Low (HML) buy-sell matrix (e.g., HH refers to the HIGH BUY, HIGH SELL cell), as follows:

DUMMY1: equals 1 if the half-hour interval falls in the LL cell

DUMMY2: equals 1 if the half-hour interval falls in the MM cell

DUMMY3: equals 1 if the half-hour interval falls in the LH or HL cells

DUMMY4: equals 1 if the half-hour interval falls in the MH or HM cells

DUMMY5: equals 1 if the half-hour interval falls in the HH cell

Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. The omitted cells are the (LM, ML) cells of the HML matrix.

In addition, HILO is regressed on the following control variables:

- Log of the number of trades in an interval
- IMBALANCE: log ratio of the absolute imbalance to the total number of trades, where imbalance is the number of buyer-initiated minus the number of seller-initiated trades
- NEWS: a dummy variable that equals 1 on days with news.
- [Open, 15 min after open]: a dummy variable that equals 1 when the trade occurs in the first 15 minutes of the trading day.
- [15 min to 30 min after open]: a dummy variable that equals 1 when the trade occurs from 15 to 30 minutes after the open.
- [30 min to 15 min before close]: a dummy variable that equals 1 when the trade occurs from 15 to 30 minutes before close.
- [15 min before close, close]: a dummy variable that equals 1 when the trade occurs in the last 15 minutes of the trading day.
- Log of the previous day's closing price
- PEBAS: the proportional effective bid-ask half-spread, equal to Q*(P- M)/M, where P is the trade price, M is the quote mid-point, and Q is +1 (-1) for a buyer (seller) initiated trade.
- 3 lags of HILO. For the first-half hour of the day, we use the absolute value of the return from the previous day's closing to the current day's opening price as the first lag of HILO.

Estimates have been multiplied by 100. *T*-statistics are corrected for autocorrelation and heteroskedasticity using the Newey-West procedure and 14 lags. A ** indicates significance at 1 per cent level or less; * indicates significance at 5 percent level or less.

Table 6 (continued)

Panel A: NYSE stocks, large and All trade sizes

	All trade sizes		Large trades	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Intercept	0.0417	1.05	0.1199**	4.20
Dummy1 (LL)	-0.0630**	-15.28	-0.0840**	-17.73
Dummy2 (MM)	0.0598**	9.27	0.0382**	7.01
Dummy3 (HL,LH)	0.0416**	7.09	0.0712**	9.61
Dummy4 (MH,HM)	0.1271**	20.56	0.1005**	15.81
Dummy5 (HH)	0.2480**	31.82	0.2209**	21.06
Log of NUMBER OF TRADES	0.1391**	45.21	0.1667**	46.44
IMBALANCE	-0.0040**	-2.85	0.0027*	2.55
[Open, 15 min after open]	0.1617**	11.80	0.1812**	13.25
[15 min to 30 min after open]	-0.1090**	-11.95	-0.0700**	-6.94
[30 min to 15 min before close]	-0.2790**	-49.28	-0.2770**	-44.77
[15 min before close, close]	-0.2870**	-39.17	-0.2730**	-35.90
News day dummy	-0.0150**	-3.21	-0.0140**	-2.81
Log of prior day closing price	-0.1190**	-13.37	-0.1650**	-24.03
PEBAS	1.3322**	13.46	0.4818**	14.47
HILO, LAG 1	16.2466**	13.74	17.5111**	13.54
HILO, LAG 2	11.0778**	13.95	12.2059**	14.08
HILO, LAG 3	9.0829**	13.69	9.7430**	13.84
Adjusted R-squared	0.49		0.46	
Number of observations	61,899		55,478	

Panel B: Nasdaq stocks, large and All trade sizes

	All trade sizes		Large trades	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Intercept	-0.4970**	-16.66	-0.3550**	-14.93
Dummy1 (LL)	-0.0720**	-15.50	-0.0880**	-17.60
Dummy2 (MM)	0.0452**	4.72	0.0595**	8.13
Dummy3 (HL,LH)	0.0316**	5.31	0.0148	1.94
Dummy4 (MH,HM)	0.1325**	17.48	0.0831**	11.82
Dummy5 (HH)	0.3403**	37.23	0.2157**	20.84
Log of NUMBER OF TRADES	0.1839**	56.72	0.1905**	59.12
IMBALANCE	-0.0020	-1.29	0.0005	0.39
[Open, 15 min after open]	0.2205**	17.80	0.3946**	29.10
[15 min to 30 min after open]	-0.2200**	-22.98	-0.1320**	-12.81
[30 min to 15 min before close]	-0.3130**	-54.00	-0.2770**	-45.18
[15 min before close, close]	-0.3240**	-37.65	-0.2050**	-24.32
News day dummy	-0.0310**	-6.39	-0.0270**	-5.12
Log of prior day closing price	-0.0260**	-4.69	-0.0740**	-14.55
PEBAS	3.8498**	31.79	2.5103**	32.85
HILO, LAG 1	15.7244**	23.28	18.0644**	24.74
HILO, LAG 2	8.7687**	15.80	10.6162**	17.55
HILO, LAG 3	6.4857**	13.22	7.4386**	13.96
Adjusted R-squared	0.55		0.51	
Number of observations	62,069		56,788	

Table 7: Regressions of Trading Costs on Sidedness and Clustering

The table shows results from a regression of the trading costs on dummy variables for sidedness and clustering. The sample is 41 NYSE (Panel A) and 41 Nasdaq stocks (Panel B) during January 2 to May 31 2003. The NYSE and Nasdaq stocks are matched according to their closing price and market value on December 31 2002. Statistics are reported separately for the sample of All trade sizes (*All*) and the sample of large trades (*Large*), defined as those in the top 10 percentile of the dollar value of trades of a stock in the sample. The proxy for trading costs is PEBAS, the average proportional effective bid-ask half-spread in a half-hour interval. PEBAS is Q*(P- M)/M, where P is the trade price, Q is +1 (-1) for a buyer (seller) initiated trade, and M is the quote mid-point. The trading cost measure is regressed on dummy variables for sidedness and clustering. They refer to cells in the 3x3 High-Medium-Low (HML) buy-sell matrix (e.g., HH refers to the HIGH BUY, HIGH SELL cell), as follows:

DUMMY1: equals 1 if the half-hour interval falls in the LL cell

DUMMY2: equals 1 if the half-hour interval falls in the MM cell

DUMMY3: equals 1 if the half-hour interval falls in the LH or HL cells

DUMMY4: equals 1 if the half-hour interval falls in the MH or HM cells

DUMMY5: equals 1 if the half-hour interval falls in the HH cell

Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. The omitted cells are the (LM, ML) cells of the HML matrix.

In addition, the trading cost measure is regressed on the following control variables:

- Log of the number of trades in a half-hour interval
- IMBALANCE: log ratio of the absolute imbalance to the total number of trades, where imbalance is the number of buyer-initiated minus the number of seller-initiated trades
- NEWS: a dummy variable that equals 1 on days with news.
- [Open, 15 min after open]: a dummy variable that equals 1 when the trade occurs in the first 15 minutes of the trading day.
- [15 min to 30 min after open]: a dummy variable that equals 1 when the trade occurs from 15 to 30 minutes after the open.
- [30 min to 15 min before close]: a dummy variable that equals 1 when the trade occurs from 15 to 30 minutes before close.
- [15 min before close, close]: a dummy variable that equals 1 when the trade occurs in the last 15 minutes of the trading day.
- Log of the previous day's closing price
- HILO: the maximum minus the minimum price in a half-hour
- 3 lags of PEBAS

Estimates have been multiplied by 100. *T*-statistics are corrected for autocorrelation and heteroskedasticity using the Newey-West procedure. A ** indicates significance at 1 per cent level or less; * indicates significance at 5 percent level or less.

Table 7 (continued)

Panel A: PEBAS, NYSE stocks

	All trade sizes		Large trades	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Intercept	0.1033**	18.35	0.2584**	30.76
Dummy1 (LL)	-0.0020**	-2.72	-0.0040	-1.70
Dummy2 (MM)	-0.0020*	-1.98	-0.0080**	-4.02
Dummy3 (HL,LH)	0.0040**	3.69	0.0036	1.08
Dummy4 (MH,HM)	0.0001	0.07	-0.0100**	-4.70
Dummy5 (HH)	-0.0010	-1.40	-0.0140**	-6.08
Log of NUMBER OF TRADES	-0.0120**	-20.16	-0.0250**	-18.70
IMBALANCE	0.0006*	2.43	0.0011**	2.75
[Open, 15 min after open]	0.0764**	40.61	0.1416**	27.34
[15 to 30 min after open]	0.0030*	2.00	0.0118**	3.21
[15 min before close, close]	0.0091**	10.40	0.0085**	3.09
[30 to 15 min before close]	0.0181**	16.26	0.0202**	9.94
News day dummy	-0.0020**	-2.94	-0.0020	-1.59
Log of prior day closing price	-0.0150**	-13.68	-0.0380**	-23.99
HILO	3.2074**	33.94	5.8128**	29.14
PEBAS LAG1	0.3281**	22.11	0.1368**	9.24
PEBAS LAG2	0.1383**	11.01	0.0758**	4.46
PEBAS LAG3	0.1299**	10.94	0.0391**	3.52
Adjusted R-squared	0.53		0.26	
Number of observations	61,896		45,173	

Panel B: PEBAS, Nasdaq stocks

	All trade sizes		Large trades	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Intercept	0.0708**	30.45	0.1083**	30.43
Dummy1 (LL)	-0.0003	-0.80	0.0003	0.56
Dummy2 (MM)	-0.0005	-0.65	-0.0008	-1.14
Dummy3 (HL,LH)	0.0014**	3.50	0.0020**	2.67
Dummy4 (MH,HM)	-0.0005	-1.14	-0.0007	-1.18
Dummy5 (HH)	-0.0030**	-8.03	-0.0030**	-4.39
Log of NUMBER OF TRADES	-0.0070**	-26.74	-0.0110**	-25.90
IMBALANCE	0.0009**	8.30	0.0002*	2.00
[Open, 15 min after open]	0.0350**	44.95	0.0288**	28.24
[15 to 30 min after open]	0.0008	1.37	0.0073**	8.95
[30 to 15 min before close]	0.0067**	17.69	0.0069**	10.17
[15 min before close, close]	0.0167**	34.20	0.0158**	21.11
News day dummy	0.0006**	2.64	0.0008*	2.06
Log of prior day closing price	-0.0100**	-31.98	-0.0140**	-28.98
HILO	1.1771**	26.25	1.3962**	21.89
PEBAS LAG1	0.4225**	42.31	0.2495**	23.45
PEBAS LAG2	0.1354**	12.99	0.1677**	17.30
PEBAS LAG3	0.1435**	17.94	0.1649**	17.23
Adjusted R-squared	0.77		0.60	
Number of observations	62,067		48,949	

Table 8: Effect of Trade Classification Errors on Sidedness and Clustering

Panel A of the table reports the distribution of buyer-initiated and seller-initiated trades. Each cell of the table reports, averaged over stocks, and for a particular buy-and-sell-trade arrival combination, the observed and **unexpected (in bold)** percent of half-hours, and the chi-square statistic of the cell as a percent of the overall chi-square. Numbers are reported for the following buy-and-sell-trade arrival combinations: low buyer-initiated and low seller-initiated trade arrivals (LL), medium buyer-initiated and medium seller-initiated trade arrivals (MM), high buyer-initiated and low seller-initiated trade arrivals (HL), low buyer-initiated and high seller-initiated trade arrivals (LH), and high buyer-initiated and high seller-initiated trade arrivals (HH). Statistics are shown for trades that are inside quotes (but not at the mid-quote), trades at the mid-quote, trades for the 20 largest stocks, and trades for the pre-decimalization period of June 2000. Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. Further details of calculations are in the text of Table 4. Results from hypotheses tests are shown under the heading, *Mean Differences in Observed Percent of Half-Hours*. The comparison is for stocks commonly traded in the two samples. We show *t*-statistics and *p*-values for the null hypotheses that the difference in mean (μ) of the observed percent of half-hours between all trades and trades inside quotes, all trades and trades at the mid-quote, the 20 largest and smallest stocks, or post- and pre-decimalization trades, is zero for (1) the diagonal cells and (2) the HL and LH cells. The standard errors used to compute the *t*-statistics are from a Poisson regression of cell counts on cell and table dummies, as described in Appendix A of the text. ** (*) indicates whether the means are significantly different, at the one (five) percent level or less.

Panel B of the table shows results from a regression of volatility and trading costs on dummy variables for sidedness and clustering. Statistics are reported separately for trades inside quotes (but not at the mid-quote), and for trades at the mid-quote. The measure of volatility is HILO, log of the highest to lowest price in an interval. The proxy for trading costs is PEBAS, the average proportional effective bid-ask half-spread in an interval, for trades inside quotes and PQBAS, the average proportional quoted bid-ask half-spread in an interval, for trades at the mid-quote. Dummy variables for sidedness and clustering refer to cells in the 3x3 High-Medium-Low (HML) buy-sell matrix (e.g., HH refers to the HIGH BUY, HIGH SELL cell), as follows:

DUMMY1: equals 1 if the half-hour interval falls in the LL cell

DUMMY2: equals 1 if the half-hour interval falls in the MM cell

DUMMY3: equals 1 if the half-hour interval falls in the LH or HL cells

DUMMY4: equals 1 if the half-hour interval falls in the MH or HM cells

DUMMY5: equals 1 if the half-hour interval falls in the HH cell

The omitted cells are the (LM, ML) cells of the HML matrix. In addition, we include the following explanatory variables:

- Log of the number of trades in a half-hour interval
- IMBALANCE: log ratio of the absolute imbalance to the total number of trades, where imbalance is the number of buyer-initiated minus the number of seller-initiated trades

Finally, control variables for NEWS, time-of-day effects, the share price, 3 lags of the dependent variable and either trading costs (when HILO is the dependent variable) or HILO (when trading cost is the dependent variable) are included. Results for the control variables are not reported to conserve space.

The sample is 41 NYSE and 41 Nasdaq stocks during January 2 to May 31 2003 for the post-decimalization sample and June 2000 for the pre-decimalization sample. The NYSE and Nasdaq stocks are matched according to their closing price and market value on December 31 2002.

Table 8 Panel A: Effect of Trade Classification Errors on Sidedness and Clustering

Distribution of buyer-initiated and seller-initiated trades

	All trade sizes, NYSE stocks					All trade sizes, Nasdaq stocks				
	H,L	L,H	H,H	L,L	M,M	H,L	L,H	H,H	L,L	M,M
Trades that are inside quotes but not at quote midpoint										
Obs %	5.23	5.25	11.65	23.29	13.33	4.28	4.36	16.94	32.25	8.82
Unexp %	**-5.21**	**-5.13**	**5.02**	**6.71**	**1.39**	**-8.83**	**-8.80**	**8.87**	**10.68**	**1.92**
Chi-sq share %	2.71	2.74	79.45	4.95	0.89	1.41	1.42	86.35	5.24	0.58
Trades that are at quote midpoint										
Obs %	5.62	6.26	5.66	24.20	16.97	4.64	4.51	8.97	24.72	17.25
Unexp %	**-1.94**	**-2.08**	**1.98**	**3.02**	**0.99**	**-3.97**	**-4.02**	**4.29**	**5.35**	**1.64**
Chi-sq share %	4.50	4.09	73.29	2.34	1.45	3.09	3.25	79.57	4.16	0.97
20 largest stocks										
Obs %	6.01	6.02	17.21	25.72	8.35	7.68	7.53	22.44	37.57	2.67
Unexp %	**-7.05**	**-7.19**	**7.51**	**7.79**	**1.06**	**-10.12**	**-10.45**	**10.43**	**10.79**	**0.65**
Chi-sq share %	2.61	2.75	72.90	10.88	1.04	4.24	3.74	67.77	14.18	0.70
Pre-decimalization period										
Obs %	4.09	4.76	12.57	21.06	15.86	6.94	7.32	15.69	33.62	5.69
Unexp %	**-5.51**	**-5.41**	**5.92**	**6.15**	**1.16**	**-7.23**	**-7.26**	**7.12**	**7.80**	**0.43**
Chi-sq share %	3.58	3.62	61.30	9.38	2.75	5.29	5.19	56.33	13.85	1.62

Mean difference in observed percent of half-hours

	All trade sizes, NYSE stocks				All trade sizes, Nasdaq stocks			
Null hypothesis	**No. stocks commonly traded**	**Estimate**	**T-statistic**	**P-value**	**No. stocks commonly traded**	**Estimate**	**T-statistic**	**P-value**
All trades – Trades inside quotes but not at quote mid-point								
μ(LL+MM+HH)=0 μ(LH+HL)=0	41	2.51** 1.87**	6.04 9.32	0.0000 0.0000	41	2.17** 6.39**	4.82 31.34	0.0000 0.0000
All trades – Trades at quote midpoint								
	41	3.53** 0.48*	8.52 2.32	0.0000 0.0207	41	7.89** 5.87**	17.79 28.23	0.0000 0.0000
20 largest stocks – 20 smallest stocks								
	----	1.60** -0.90**	2.62 -2.94	0.0091 0.0034	----	4.88** 0.37	7.37 1.13	0.0000 0.2596
Post-decimalization – pre-decimalization								
μ(LL+MM+HH)=0 μ(LH+HL)=0	41	1.29 3.50**	1.84 11.25	0.0667 0.0000	41	5.18** 0.77*	6.94 2.02	0.0000 0.0440

Table 8 Panel B: Effect of trade classification errors on volatility and trading cost regressions

Dependent variable: HILO

	All trade sizes, NYSE		All trade sizes, Nasdaq	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Trades that are inside quotes but not at mid-quote				
Intercept	0.3967**	16.22	-0.0520*	-2.21
Dummy1 (LL)	-0.0420**	-10.81	-0.0620**	-13.69
Dummy2 (MM)	0.0689**	13.30	0.0533**	7.32
Dummy3 (HL,LH)	0.0145**	2.61	0.0151*	2.21
Dummy4 (MH,HM)	0.1393**	24.43	0.1387**	19.67
Dummy5 (HH)	0.2959**	33.61	0.3358**	37.67
Log of NUMBER OF TRADES	0.1629**	55.52	0.1874**	61.02
IMBALANCE	-0.0050**	-4.48	0.0004	0.24
Adjusted R-squared	0.47		0.55	
Number of observations	57,199		60,178	
Trades that are at mid-quote				
Intercept	0.1937**	12.01	-0.0870**	-3.52
Dummy1 (LL)	-0.0360**	-5.87	-0.0930**	-18.16
Dummy2 (MM)	0.0716**	11.37	0.0898**	13.85
Dummy3 (HL,LH)	-0.0090	-1.52	-0.0180*	-2.45
Dummy4 (MH,HM)	0.1078**	16.33	0.1581**	22.67
Dummy5 (HH)	0.1587**	12.13	0.3526**	27.44
Log of NUMBER OF TRADES	0.1313**	35.64	0.1698**	65.35
IMBALANCE	-0.0060**	-4.65	-0.0050**	-3.56
Adjusted R-squared	0.36		0.47	
Number of observations	39,276		45,290	

Dependent variable: PEBAS or PQBAS

	All trade sizes, NYSE		All trade sizes, Nasdaq	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Trades that are inside quotes but not at mid-quote: Dependent variable is PEBAS				
Intercept	0.0911**	8.04	0.0514**	28.05
Dummy1 (LL)	-0.0230**	-2.97	-0.0030**	-9.31
Dummy2 (MM)	-0.0290**	-3.56	0.0013*	2.39
Dummy3 (HL,LH)	0.0268**	3.33	0.0056**	9.98
Dummy4 (MH,HM)	-0.0030	-0.49	0.0022**	4.86
Dummy5 (HH)	-0.0020	-0.30	0.0004	0.88
Log of NUMBER OF TRADES	-0.0280**	-9.14	-0.0060**	-23.26
IMBALANCE	0.0164**	8.12	0.0007**	6.08
Adjusted R-squared	0.56		0.67	
Number of observations	56,960		60,157	
Trades that are at mid-quote: Dependent variable is PQBAS				
Intercept	0.1937**	13.56	0.0636**	17.70
Dummy1 (LL)	-0.0130**	-4.51	-0.0020**	-4.46
Dummy2 (MM)	0.0096*	2.53	-0.0006	-1.44
Dummy3 (HL,LH)	0.0035	0.84	0.0004	0.65
Dummy4 (MH,HM)	0.0117**	3.51	-0.0009*	-2.22
Dummy5 (HH)	0.0112**	2.99	-0.0030**	-5.30
Log of NUMBER OF TRADES	-0.0190**	-8.86	-0.0040**	-14.35
IMBALANCE	0.0013*	2.19	-0.0004**	-3.71
Adjusted R-squared	0.20		0.85	
Number of observations	39,007		45,058	

Table 9: Sidedness and Clustering on Days With Corporate News Events, For First 15 Minutes of News and Non-News Days

Panel A of the table reports descriptive statistics for news and non-news days, where news days are identified by corporate news in major publications relating to earnings, dividends, mergers and acquisitions, share buybacks or stock splits, or changes in credit ratings. The reported statistics are ACLOP, absolute excess returns from the previous day's close to the current day's open, HILO, log of the highest to lowest price in an interval, VOL, the trading volume, #TR, the number of trades and PEBAS, , the average proportional effective bid-ask half-spread in an interval. Panel B of the table shows the distribution of buyer and seller-initiated trades for the first 15 minutes of news days. Each cell of the table reports, averaged over stocks, and for a particular buy-and-sell-trade arrival combination, the observed and **unexpected (in bold)** percent of half-hours, and the chi-square statistic of the cell as a percent of the overall chi-square. Further details of calculations are in the text of Table 4. Numbers are reported for the following trade arrival combinations: low buyer-initiated and low seller-initiated trade arrivals (LL), medium buyer-initiated and medium seller-initiated trade arrivals (MM), high buyer-initiated and low seller-initiated trade arrivals (HL), low buyer-initiated and high seller-initiated trade arrivals (LH), and high buyer-initiated and high seller-initiated trade arrivals (HH). Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. Panel C shows *t*-statistics and *p*-values for the null hypothesis that the mean (μ) difference in the observed percent of 15-minute intervals between the first 15 minutes of news days and all days is zero for (1) the diagonal cells and (2) the HL and LH cells. The comparison is for stocks commonly traded in the two samples. The standard errors used to compute the *t*-statistics are obtained from a Poisson regression of cell counts on cell and table dummies, as described in Appendix A of the text. ** (*) indicates whether the means are significantly different, at the one (five) percent level or less.

Panel D of the table shows results from a regression of volatility and trading costs on dummy variables for sidedness and clustering. The measure of volatility is HILO. The proxy for trading costs is PEBAS. The dummy variables for sidedness and clustering refer to cells in the 3x3 High-Medium-Low (HML) buy-sell matrix (e.g., HH refers to the HIGH BUY, HIGH SELL cell), as follows:

DUMMY1: equals 1 if the half-hour interval falls in the LL cell

DUMMY2: equals 1 if the half-hour interval falls in the MM cell

DUMMY3: equals 1 if the half-hour interval falls in the LH or HL cells

DUMMY4: equals 1 if the half-hour interval falls in the MH or HM cells

DUMMY5: equals 1 if the half-hour interval falls in the HH cell

The omitted cells are the (LM, ML) cells of the HML matrix. In addition, we include the following explanatory variables:

- Log of the number of trades in a half-hour interval
- IMBALANCE: log ratio of the absolute imbalance to the total number of trades, where imbalance is the number of buyer-initiated minus the number of seller-initiated trades
- NEWS: a dummy variable that equals 1 on news days and is zero otherwise.

Finally, control variables for time-of-day effects, the share price, 3 lags of the dependent variable and either trading costs (when HILO is the dependent variable) or HILO (when trading cost is the dependent variable) are included. Results for the control variables are not reported to conserve space.

The sample is 41 NYSE and 41 Nasdaq stocks during January 2 to May 31 2003. The NYSE and Nasdaq stocks are matched according to their closing price and market value on December 31 2002.

Table 9: Sidedness and clustering on days with corporate news events

Panel A: Descriptive statistics for news and non-news days

	Obs	Aclop	HILO	VOL	#TR	Pebas	Obs	Aclop	HILO	VOL	#TR	Pebas
News days	4,431	1.09**	0.90**	226,585*	109**	0.067**	3,184	1.50**	1.06**	398,013**	275**	0.067
No-news days	49,795	0.74	0.75	114,223	88	0.063	51,231	0.92	0.89	268,023	195	0.063

Panel B: Distribution of buyer-initiated and seller-initiated trades

	All trade sizes, NYSE stocks					All trade sizes, Nasdaq stocks				
	H,L	**L,H**	**H,H**	**L,L**	**M,M**	**H,L**	**L,H**	**H,H**	**L,L**	**M,M**
First 15 minutes of news days	5.15	4.40	13.40	18.33	22.19	7.55	7.84	16.34	30.39	14.20
	-5.86	**-5.50**	**5.03**	**4.73**	**-1.60**	**-6.74**	**-6.51**	**6.51**	**7.44**	**0.69**
	9.25	8.71	15.15	18.21	8.63	9.35	8.66	19.75	29.41	7.44

Panel C: Mean difference in observed percent of half-hours, for first 15 minutes of all days and news days

	All trade sizes, NYSE stocks				All trade sizes, Nasdaq stocks			
Null hypothesis	**No. stocks common**	**Estimate**	**T-statistic**	**P-value**	**No. stocks common**	**Estimate**	**T-statistic**	**P-value**
First 15 minutes, all days –First 15 minutes, news days								
μ(LL+MM+HH)=0	39	-5.29	-1.31	0.1905	28	-15.12**	-2.63	0.0087
μ(LH+HL)=0		1.65	0.72	0.4727		4.16	1.62	0.1051

Panel D: Effect of sidedness and clustering on volatility and trading costs

	All trade sizes, NYSE		All trade sizes, Nasdaq	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Dependent variables is HILO				
NEWS DAY DUMMY	-0.0140	-1.52	0.0072	0.55
Dummy1 (LL)	-0.0650**	-15.70	-0.0720**	-15.65
Dummy2 (MM)	0.0611**	9.50	0.0473**	4.97
Dummy3 (HL,LH)	0.0447**	7.64	0.0342**	5.75
Dummy4 (MH,HM)	0.1299**	21.06	0.1352**	17.86
Dummy5 (HH)	0.2497**	32.19	0.3345**	36.81
Log of NUMBER OF TRADES	0.1362**	43.13	0.1806**	56.26
IMBALANCE	-0.0040**	-2.94	-0.0020	-1.27
Adjusted R-squared	0.49		0.55	
Number of observations	61,899		62,069	
Dependent variables is PEBAS				
NEWS DAY DUMMY	-0.0005	-0.54	0.0007	1.66
Dummy1 (LL)	-0.0020**	-2.70	-0.0003	-0.85
Dummy2 (MM)	-0.0020*	-2.00	-0.0005	-0.64
Dummy3 (HL,LH)	0.0040**	3.69	0.0014**	3.51
Dummy4 (MH,HM)	0.0000	0.05	-0.0005	-1.10
Dummy5 (HH)	-0.0020	-1.48	-0.0030**	-7.95
Log of NUMBER OF TRADES	-0.0120**	-20.12	-0.0070**	-26.79
IMBALANCE	0.0006*	2.42	0.0009**	8.31
Adjusted R-squared	0.53		0.77	
Number of observations	61,896		62,067	

Table 10: Sidedness and Clustering Using One-Minute Windows, For First 15 Minutes of Trading Days

The table reports the distribution of buyer and seller-initiated trades, and their association with volatility and trading costs, for 1-minute windows for the first 15 minutes of trading days. Panel A shows the distribution of buyer and seller-initiated trades for the first 15 minutes for all days and news days. News days for a stock are the 30 percentile of days with the largest values of ACLOP, the absolute value of the log excess returns from the previous day's closing price to the current day's opening price. Each cell of the table reports, averaged over stocks, and for a particular buy-and-sell-trade arrival combination, the observed and **unexpected (in bold)** percent of half-hours, and the chi-square statistic of the cell as a percent of the overall chi-square. Further details of calculations are in the text of Table 4. Numbers are reported for the following trade arrival combinations: low buyer-initiated and low seller-initiated trade arrivals (LL), medium buyer-initiated and medium seller-initiated trade arrivals (MM), high buyer-initiated and low seller-initiated trade arrivals (HL), low buyer-initiated and high seller-initiated trade arrivals (LH), and high buyer-initiated and high seller-initiated trade arrivals (HH). Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. Panel B shows *t*-statistics and *p*-values for the null hypothesis that the mean (μ) difference in the observed percent of 15-minute intervals between the first 15 minutes of news days and all days is zero for (1) the diagonal cells and (2) the HL and LH cells. The comparison is for stocks commonly traded in the two samples. The standard errors used to compute the *t*-statistics are obtained from a Poisson regression of cell counts on cell and table dummies, as described in Appendix A of the text. ** (*) indicates whether the means are significantly different, at the one (five) percent level or less.

Panel C of the table shows results from a regression of volatility and trading costs on dummy variables for sidedness and clustering. The measure of volatility is HILO. The proxy for trading costs is PEBAS. The dummy variables for sidedness and clustering refer to cells in the 3x3 High-Medium-Low (HML) buy-sell matrix (e.g., HH refers to the HIGH BUY, HIGH SELL cell), as follows:

DUMMY1: equals 1 if the half-hour interval falls in the LL cell

DUMMY2: equals 1 if the half-hour interval falls in the MM cell

DUMMY3: equals 1 if the half-hour interval falls in the LH or HL cells

DUMMY4: equals 1 if the half-hour interval falls in the MH or HM cells

DUMMY5: equals 1 if the half-hour interval falls in the HH cell

The omitted cells are the (LM, ML) cells of the HML matrix. In addition, we include the following explanatory variables:

- Log of the number of trades in a half-hour interval
- IMBALANCE: log ratio of the absolute imbalance to the total number of trades, where imbalance is the number of buyer-initiated minus the number of seller-initiated trades

Finally, control variables for NEWS, the share price, 3 lags of the dependent variable and either trading costs (when HILO is the dependent variable) or HILO (when trading cost is the dependent variable) are included. Results for the control variables are not reported to conserve space.

The sample is 41 NYSE and 41 Nasdaq stocks during January 2 to May 31 2003. The NYSE and Nasdaq stocks are matched according to their closing price and market value on December 31 2002.

Table 10 (continued)

Panel A: Distribution of buyer-initiated and seller-initiated trades

	All trade sizes, NYSE stocks					All trade sizes, Nasdaq stocks				
	H,L	L,H	H,H	L,L	M,M	H,L	L,H	H,H	L,L	M,M
First 15 minutes, all days	7.89	6.97	4.93	34.11	11.98	7.53	7.41	10.02	36.13	6.08
	-2.24	**-2.08**	**1.10**	**5.00**	**1.78**	**-4.79**	**-4.73**	**4.64**	**5.42**	**0.54**
	4.27	4.20	63.63	8.94	1.86	1.75	1.60	90.37	1.21	0.29
First 15 minutes of news days	7.46	6.69	4.03	4.97	21.35	8.27	8.21	10.21	27.40	6.86
	0.95	**1.22**	**0.39**	**-5.43**	**-2.87**	**-3.90**	**-3.71**	**4.22**	**2.53**	**-0.85**
	5.47	7.24	46.10	9.44	4.33	4.50	4.06	67.56	5.33	1.64

Panel B: Mean difference in observed percent of half-hours, for first 15 minutes of all days and news days

	All trade sizes, NYSE stocks				All trade sizes, Nasdaq stocks			
Null hypothesis	**No. stocks common**	**Estimate**	**T-statistic**	**P-value**	**No. stocks common**	**Estimate**	**T-statistic**	**P-value**
First 15 minutes, all days –First 15 minutes, news days								
μ(LL+MM+HH)=0	41	19.09**	31.80	0.0000	41	5.25**	8.76	0.0000
μ(LH+HL)=0		1.51**	4.14	0.0000		-1.94**	-5.59	0.0000

Panel C: Effect of sidedness and clustering on volatility and trading costs, for first 15 minutes of trading days

	All trade sizes, NYSE		All trade sizes, Nasdaq	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Dependent variables is HILO				
Intercept	0.1971**	17.43	-0.0840**	-4.41
Dummy1 (LL)	0.0017	0.24	-0.0060*	-2.19
Dummy2 (MM)	0.0631**	12.96	0.0573**	13.42
Dummy3 (HL,LH)	0.0134*	2.48	0.0205**	5.36
Dummy4 (MH,HM)	0.0900**	15.00	0.1065**	22.76
Dummy5 (HH)	0.1207**	11.65	0.1825**	26.13
Log of NUMBER OF TRADES	0.0843**	21.63	0.1060**	64.91
IMBALANCE	-0.0030**	-2.97	-0.0080**	-9.82
Adjusted R-squared	0.28		0.55	
Number of observations	26,663		46,228	
Dependent variables is PEBAS				
Intercept	0.2472**	6.90	0.1010**	29.60
Dummy1 (LL)	0.0090	0.53	-0.0070**	-9.93
Dummy2 (MM)	-0.0170*	-2.27	0.0009	0.88
Dummy3 (HL,LH)	0.0515**	4.16	0.0085**	10.04
Dummy4 (MH,HM)	-0.0190*	-2.52	0.0076**	8.89
Dummy5 (HH)	-0.0030	-0.25	0.0222**	21.29
Log of NUMBER OF TRADES	-0.0080	-1.33	-0.0160**	-40.03
IMBALANCE	0.0052**	3.22	0.0025**	13.12
Adjusted R-squared	0.10		0.58	
Number of observations	25,519		45,604	

Figure 1: Distribution of buyer and seller-initiated trades

The figures illustrate the Pearson chi-square statistic, as a percent of the total chi-square, for three combinations of buyer-initiated and seller-initiated trades. The combinations are 1=LOW, 2=MEDIUM and 3=HIGH. The LOW, MEDIUM, HIGH trade arrivals are determined relative to what would be expected if buyer-initiated and seller-initiated trades follow Poisson arrival processes.