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

We thank Markus Brunnermeier, Joel Hasbrouck, Terry Hendershott, Charles Jones, Albert Menkveld, Lasse Pederson and Jiang Wang, and seminar participants at the AFA 2006 meetings, the NBER Market Microstructure conference of October 2005, the Microstructure conference in Norges Bank (Oslo), Baruch College, Rice University, Rutgers University, SUNY Binghamton, and the Federal Reserve Bank of New York, 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 two-sided and that trades cluster in certain half-hour periods 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 in half-hour intervals are positively correlated; by "trade clustering" we mean that trades tend to bunch together in certain half-hour intervals with greater frequency than would be expected if their arrival was a random process. Controlling for trading volume, news, and other microstructure effects, we find that two-sided clustering leads to 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 half-hour intervals ("the sidedness of markets" in our terminology), and examine its implications for price volatility and trading costs, for a sample of NYSE and Nasdaq stocks. Our primary motivation is that the sidedness of markets both sheds light on the price formation process and reflects the trading motives of participants.

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 private signals or on different interpretations of public news, leads to trade clustering on both sides of the market, with elevated volatility during these periods. Finally, trades to rebalance portfolios also imply two-sided markets, but do not imply a relation between sidedness, volatility, and trading costs. We discuss these motives in further detail, and present empirical findings that underscore the importance of the second, which involves the divergence of expectations and differential private signals.[1] We stress, however, that the validity of one motive need not be at the expense of another, and that all three are likely to be important to varying degrees in understanding the dynamics of trading in financial markets.

We find that, for practically every stock in our sample, the arrivals of buyer-initiated and seller-initiated trades within half-hour intervals are positively correlated (i.e., that buyers trigger more trades in a half-hour interval when a larger number of sellers have triggered trades in that same interval, and vice versa.). In our terminology, markets are *two-sided*. In addition, buyer-initiated and seller-initiated trades tend to cluster together in particular half-hour intervals. By clustering, we mean that trading frequency is distributed in half-hour intervals such that periods with unusually high numbers of trades are more prevalent than would be expected under random trade arrival. We find that trade clustering occurs under a broad range of market settings – news

[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.

and non-news days, different times of the day, alternative market structures, and a spectrum of trade sizes. We observe that the clustering occurs for buyer-initiated and seller-initiated trades simultaneously, which means that trade clusters are two-sided.

After establishing the prevalence of two-sided trade clustering, we use regression analysis to examine the impact that two-sided trade bursts have on price volatility (measured by hi-low price ranges) and trading costs (measured by the bid-ask spread), after controlling for order imbalance, trading frequency, time-of-day effects, the arrival of news and the share price. We find that the sidedness of trades is significantly correlated with volatility and trading costs. Specifically, volatility is highest during half-hour intervals with a 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, and vice versa), after accounting for volatility. Our results imply that empirical studies of the determinants of volatility and trading costs need to take account of the sidedness of markets.

We conduct a number of additional investigations. We examine whether markets for smaller stocks are more one-sided than for larger stocks due, for example, to a lack of sufficient investor interest. We further address the concern that our results are peculiar to the post-decimalization period when trade sizes decreased substantially and large orders were more likely to have been broken up, by repeating the analysis for the pre-decimalization period of June 2000. Next, we use alternative measures of news arrival and volatility. Finally, we examine whether trade clustering may be due to the presence of stale limit orders. In all cases, we show that markets exhibit two-sided clustering and, further, volatility is highest in such markets whereas trading costs are highest in one-sided markets.

Our analysis is related to the literature that uses the autoregressive conditional duration (ACD) method to model inter-trade arrival times (for example, 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 (i.e., a half-hour). 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 half-hours and that volatility and trading costs tend to be highest in these half-hours. 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 also consider the implications for volatility and trading costs.[2] Further, whereas ACD analysis is typically stock-specific, our methodology allows us to aggregate across stocks and thus to compare aggregate trade clustering for different conditions (e.g. between days with and without news). On the other hand, ACD models allow for a richer dynamic structure in duration and for the interaction of duration and price formation.[3]

Easley, Kiefer and O'Hara (1997) and Easley, Engle, O'Hara, and Wu (2005) analyze the arrival rates of informed and uninformed trades. Based on a model where some traders have superior information compared to others, the papers investigate the dynamics of trades and trade composition. Specifically, they use data on the daily number of buyer-initiated and seller-initiated trades to estimate the probability of informed trades having occurred. Our results characterize empirically the intra-day distribution of buyer-initiated and seller-initiated trades that underlie the estimated daily probability of informed trading. We show the existence of a positive correlation between buyer-initiated and seller-initiated trades and argue that the evidence is consistent with models of differential private information or divergent beliefs about public news.

The inclusion of market imbalance and trading volume, along with sidedness, in the regression tests may be viewed as a robustness check, in that these three variables can each capture different aspects of similar events. Half-hour intervals that show high levels of two-sided clustering are no doubt also characterized by high trading volume, but not necessarily by sizable order imbalance; half-hour intervals that are one-sided are no doubt also characterized by sizable order imbalance but not necessarily by high trading volume. Do all three variables have meaningful, independent relationships with our independent variables, price volatility and trading costs? The regression results show that sidedness and trading volume are significantly determinants of volatility and trading costs, consistent with past research, while the imbalance variable yields conflicting and statistically less significant results.

[2] Hautsch (2001) examines the waiting time for a given amount of volume to be either bought or sold, or to be traded on both sides of the market simultaneously. However, he does not consider the joint arrivals of buyer-initiated and seller-initiated trades.

[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.

There has been interest in the role of order imbalance and trading volume in the price formation process. Chordia, Roll and Subrahmanyam (2002) show that daily order imbalances are negatively correlated with liquidity. In comparison, our paper focuses on the joint distribution of the number of buyer-initiated and seller-initiated trades at the intra-day frequency, and on the implications for volatility and trading costs after controlling for order imbalance. There is also a large literature that considers the effect of a given market setting (e.g., the time of day or release of news) on trading volume and order imbalance. In contrast, we examine the trade arrival process for a given market setting – e.g., whether, during the opening fifteen minutes, the trading frequency is unusually high on certain days relative to what would be expected for the opening interval if trade arrivals were a random process.

Our paper is organized as follows. In Section 1, we provide a conceptual framework for understanding our results in terms of information processes and price formation. 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 and, respectively, price volatility and trading costs. In section 6, we examine whether are results are different for small and large stocks, for the pre-decimalization period, for alternative definitions of news and volatility, and examine the role of stale limit orders in trade clustering. 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. Information and Price Formation: Alternative Views

The nature of price formation is a central topic in the microstructure literature. A key idea in the literature is that investors' trading motives may be inferred from trading volume, the sign of trades and the duration between trades. 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 than others) or heterogeneous beliefs (i.e. investors have different interpretations of public 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, based on dynamic trading models, and to what extent our findings support the predictions. 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. 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.[5] 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.[6] 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, 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 observed on both sides of the market.[7] 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

---

[4] We note that one-sided order flow would not obtain in models where price changes follow a martingale (e.g. Kyle, 1985). If the price change is proportional to order flow (with a fixed constant of proportionality), then order flow must also be a martingale.

[5] Numerical solutions in Foster and Viswanathan (1993) also suggest the possibility of trade clustering in early and late periods after the arrival of information.

[6] 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 informed investor 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.

[7] 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.

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 would tend to increase 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.[8] 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).[9] 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, rebalancing trades do not generate additional volatility or trading costs (He and Wang, 1995).

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 (investors have different private information signals) or model 2B (investors have

---

[8] 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. Thus, it appears that speculative trades in this model tend to be buyer-initiated, resulting in a one-sided market.
[9] Hendershott and Jones (2005) and Antunovich and Sarkar (2005) provide empirical evidence on order flow externality.

different interpretations of public news).[10] Two-sided markets are also predicted by model 3 (investors trade to rebalance their portfolios), but our findings on volatility and trading costs are inconsistent with model 3's predictions. In addition, because of the short (half-hour) 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.[11] Finally, 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). However, two-sided clustering obtains even when we separately examine a sample of large trades, which are less likely to be from liquidity traders.

The findings of two-sided markets, associated with high volatility, 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.

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.[12] 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.[13]

---

[10] Frankel and Froot (1990) also find a positive association between dispersion and price volatility.

[11] Informal conversations with practitioners have revealed that index funds, for instance, do not rebalance with undue frequency (e.g., every half hour).

[12] 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, VWAP, 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.

[13] 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, which peaked at 2303 shares in 1988 and has declined ever since, stood at 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

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.

We examine buy and sell arrivals during half-hour intervals. 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. We define large trades, for a stock, as

---

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).

[14] We initially matched 50 Nasdaq stocks with 50 NYSE stocks but lost 9 stocks in each market either because of delistings or mergers. 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, let x_j be the data for NYSE stock x and the j^{th} matching variable, and y_j be the data for Nasdaq firm y and the j^{th} matching variable, where j=1 (the market value), or 2 (the closing price). The Euclidean distance between NYSE firm x and Nasdaq firm y is:

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

those that are in the top decile of the dollar value of trades for that 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. 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 seller-initiated trades, for different times of the day, and for days with and without news. 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 times 100, and (2) HILO, which is the log of the ratio of the maximum to the minimum price in a half-hour or 15 minute period, times 100. 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 the news days, we select the 30 percentile of days where the value of ACLOP is largest.[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 on news days. For both markets, there are more trades, (including more large buy-triggered and large sell-triggered trades), and greater volume as compared to non-news days.

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 computation.

[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. Becker and Angstadt have estimated that, in 1994, institutional trading accounted for between 75% and 80% of NYSE 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.

[16] We also use alternative ways of identifying news days, such as days with macroeconomic news announcements, or days with large changes in the market index. With these alternative definitions, we found little difference in volatility between "non-news" and "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 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. To the extent that trades are attributable to news and momentum trading, one would expect the arrivals to be one-sided. Alternatively, trading based on differential interpretations of information signals is likely to be two-sided. Distinguishing between the two scenarios is important, because the implications for price formation are quite different. We further examine the extent to which buyer-initiated or seller-initiated trades cluster in particular half-hour intervals. Clustering is defined as an unusually high number of trades arriving in particular half-hour intervals (see Section 3*B* for further explanation). 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.

The larger, more predominantly institutional 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

---

[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.

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 (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.

A final reason to study large trades separately is as follows. In the sample of all trades, two-sided markets may obtain when investors trade on superior information by posting market orders and aggressive limit orders, which attract market orders from the opposite side. This scenario is less likely when we consider large trades exclusively, since the market orders on the opposite side are uninformative and unlikely to be large orders.

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 used to study sidedness in Section 3*A*. Results for individual stocks are reported in Section 3*B*, and results for the aggregate of stocks are given 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 algorithm cannot classify some trades, in particular those executed at the opening auction of the NYSE, and these are omitted from our sample.[18]

We tabulate the number of buyer-initiated and seller-initiated trades in each half-hour interval of each trading day, and record the number of half-hour windows for which each specific combination of buyer-initiated and seller-initiated trades (e.g., two large buy triggered trades and three large sell triggered trades in a half-hour window) was observed. The results are recorded in a matrix (BSELL matrix from here on). Our null hypothesis is that the large buy and the large sell arrivals are not associated. Given our large sample size (i.e., the large number of half-hours), the test statistic should be distributed approximately as chi-square if the null hypothesis is true. To test the hypothesis, we use the Pearson chi-square statistic, which reflects the differences between the observed and expected frequencies, where the expected frequencies are computed under the null hypothesis of independence. The chi-square statistic is computed as:

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

where $\varepsilon_{ij} = \frac{n_{i.} n_{j.}}{n}$, $n_{i.} = \sum_j n_{ij}$ 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, where R is the number of rows and C is the number of columns. For large values of Q_P , this test rejects the null hypothesis in favor of the alternative hypothesis that dependence exists between the buy and sell arrivals.

We consolidate the BSELL matrix across stocks so as to make statements about the aggregate of large buy and large 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, 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

[18] Lee and Radhakrishna (2000) find that the Lee-Ready classification of buys and sells is reasonably accurate; however, it will inevitably misclassify some trades which will create measurement error in our data.

the mean number of large sell trades in the sample.[19] The BSELL matrix is mapped for each stock into a 3-by-3, High-Medium-Low matrix (HML matrix from now on).

We base the mapping rule on the parameter λ of the Poisson arrival process. Specifically, a half-hour with n_b large 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.

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 used to determine the HIGH, MEDIUM, and LOW cutoffs are specific to each of our samples. For example, when computing results for the first 15 minutes, λ is the mean number of trades in the first 15 minutes of the trading day.

Half-hours with n_s large sell trades are similarly mapped into LOW, MEDIUM or HIGH SELL cells based on λ_s. 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 frequencies of large buy and sell trade arrivals, 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 for a stock the percent of observed and unexpected half-hours with, for instance, LOW BUY and LOW SELL arrivals, we sum o_{ij} and u_{ij}, respectively, over all cells i,j of the BSELL

[19] 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.

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 BUY and LOW SELL arrivals for a stock, 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 various hypotheses regarding the difference in cell means across different HML tables (e.g. comparing the mean for a particular cell between the table for all trades and the table for large trades). To obtain the standard errors of the cell means, we assume that the cell counts follow a Poisson distribution, and we estimate a Poisson regression model. Further details on the Poisson regression are given in the appendix.

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 arrival of buyer-initiated and seller-initiated trades in half-hour intervals are statistically independent for individual firms, we compute, for each stock, the Pearson chi-square statistic (details are in the appendix).[20] 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 .01 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 the row and column entries of the BSELL matrix (which implies the correlation between buyer-initiated and seller-initiated trades) for each stock, and column 6 is the P-value for testing the null hypothesis that the rank correlation coefficient is zero. For all 82 stocks, the null hypothesis of zero correlation is rejected at a high level of significance. For all of the 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

---

[20] 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.

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-initiated and seller-initiated trades indicates that markets are two-sided. The typical pattern of two-sidedness is illustrated in Figure 1 for three stocks in the NYSE sample and three in the Nasdaq sample. The stocks are those with the highest median trading frequency, an intermediate trading frequency, and the lowest trading frequency in each market, respectively. The height of each figure shows the percentage contribution of each cell in the 3X3 HML matrix to the overall chi-square of the matrix. "1" indicates the LOW row or column, "2" indicates the MEDIUM row or column, and "3" indicates the HIGH row or column of the HML matrix. The patterns portrayed in Figure 1 for the most active NYSE stock, General Motors, and the most active Nasdaq stock, Qualcomm, are broadly representative of our sample. The major contributing cell to the overall chi-square is the HH cell of the HML matrix. Figure 1 reflects similar patterns for the less active stocks. Recall that the HH cell includes half-hour 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. Aggregate 3x3 results

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 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 and LH cells. Half-hours that are neither clearly one-sided nor two-sided are in the two ML and the two HM 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 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 below each panel.

The results in the 3-by-3 tables 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 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 below the level expected for a Poisson arrival process. For example, in the HL cell, the three percentage 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 half-hour 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 orders results in smaller trades being triggered by institutional orders, and these 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 in the all half-hour sample 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 half-hours for all diagonal cells and negative unexpected percent of half-hours 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.

The final major row in Table 4 shows the clustering pattern for the first 15 minutes of days with news (as defined in Section 1) for the NYSE and the NASDAQ samples, respectively. 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 10.89, 4.89 and 47.90, and in Panel B (NASDAQ stocks), the entries are 16.32, 10.25 and 46.08. In general, the frequency of half-hours in the LH and HL off-diagonal cells is less than expected, and the chi-square contributed collectively by the diagonal cells is 40% 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 large buyer-initiated and large seller-initiated trades is substantially greater than would be expected under a Poisson arrival process.

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. Hypotheses tests

While two-sidedness seems to be the norm in a qualitative sense, the degree of two-sidedness may differ by time of day and arrival of news. 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 two samples (e.g. the all trades and the large trades 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 for the tests of hypotheses shown in the bottom part of each panel indicate statistically significant differences in the two-sided pattern for the different samples. The table for "different times of the day" shows that, for both NYSE and Nasdaq stocks, the mean percent of half-hours in the diagonal cells is larger for the whole sample relative to the first 15 minutes and, conversely, that the mean percent of half-hours in the off-diagonal cells (LH and HL) is lower. In addition, markets are more one-sided during the last 15 minutes of the trading day compared to the whole sample. 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 trades in the middle half-hours.

Next, turning to the table for "news versus non-news days," we observe that the difference in the mean percent of half-hours is not statistically significant in any instance in either the NYSE or Nasdaq markets. Thus, there is no evidence that the degree of two-sidedness is different between news days and non-news days. This result further reinforces our earlier finding that the arrival of news does not significantly alter the pattern of buyer-initiated and seller-initiated trade arrivals.

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

We conclude that trade bursts are typically two-sided 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 bursts occur for stocks assessed individually, this finding is not an artifact of pooling firms that trade a lot with firms that trade infrequently.

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 is a further determinant of price volatility. We first describe the regression methodology used to assess these relationships, and then present our findings for trades classified as either buyer-initiated or seller-initiated.

A. Regression Methodology

We use regression analysis to examine the relationships between volatility, sidedness and trade clustering, after controlling for microstructure effects, news, and time-of-day effects. Specifically, we regress HILO (the log difference between the maximum and the minimum price in a half-hour or a 15-minute interval) on dummy variables that reflect the degree of sidedness and trade clustering. Specifically, we include five dummy variables that refer to cells in the 3x3 High-Medium-Low (HML) buy-sell 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 in this analysis 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. DUMMY5 represents cells with the highest degree of two-sided clustering. 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). As reflected in Table 1, we expect volatility to be highest when markets are two-sided. Volatility is also expected to be higher in half-hour windows characterized by heavy trading clustering. Accordingly, we expect the coefficient on dummy 1 to be negative, and the other dummy coefficients to be positive, with dummies 4 and 5 having the highest coefficients.

In addition to trade clustering and sidedness, the dummy variables also capture changes in aggregate trading activity and imbalance in the relative numbers of buyer and seller-initiated trades. For example, aggregate trading activity is lower in the LL cells compared to the HH cells; and, controlling for aggregate trading activity, the imbalance between buyer and seller-initiated trades is greater in the off-diagonal cells than in the diagonal cells. To separate out the effects of trading activity and order imbalance, we include the following variables:

- Log of the number of trades in a half-hour interval
- IMBALANCE: log of the absolute value of buyer-initiated minus seller-initiated trades in a half-hour, as a proportion of 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 indicate that volatility is higher on days with news. The statistics also show strong time-of-day effects on volatility. Accordingly, we include the following variables to control for these effects:

- NEWS: a dummy variable that equals 1 on days with news (as defined in Section 1).
- [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 30 to 15 minutes before the close.
- [15 min before close]: a dummy variable that equals 1 when the trade occurs in the last 15 minutes of the trading day.

Periods with higher trading costs are likely to experience greater volatility.[21] Prior research shows that the price level is a determinant of liquidity and, since liquid stocks may be less volatile, stocks with higher prices may be less volatile. Finally, there is ample evidence that volatility is persistent. We include the following variables to control for these effects:

- Log of the previous day's closing price
- PEBAS: the proportional effective bid-ask half-spread for the interval
- Three lags of HILO.[22]

[21] See, for example, Subrahmanyam, A., 1994, Circuit breakers and market volatility: A theoretical perspective, Journal of Finance 49,237-254.
[22] 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.

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 are shown in Table 5 (Panel A for NYSE stocks, and Panel B for Nasdaq stocks). We find, for both markets, and all trade sizes, that both the mean and median volatility are generally increasing as we progress from half-hour intervals with few trades (the LL cells) to intervals with extreme one-sided trades (the LH and HL cells). Both averages are at a maximum in intervals with a large number of trades on both sides of the market (the HH cells). For example, for the sample of all trades for NYSE stocks, the median volatility 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. The 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 those of either DUMMY2 or DUMMY5. Thus, volatility is high when markets are extremely one-sided, but not as high as in moderately two-sided markets. Finally, the DUMMY4 coefficient (representing the MH and HM cells) are always positive and significant, and with magnitudes lower only than that of the DUMMY5 coefficient. Thus, markets in these half-hours appear to be more like two-sided than one-sided markets.

The above results obtain even after controlling for aggregate trading activity and imbalance. We find that volatility is significantly and positively correlated with the number of trades in both exchanges and for all trade sizes. This result is consistent with Jones, Kaul and

Lipson (1994) and Chan and Fong (2000) who show, using daily data, that volatility is positively related to trading frequency. The imbalance variable gives a mixed performance; it's coefficient is significantly positive for large NYSE trades, significantly negative for all NYSE trades, insignificantly positive for large Nasdaq trades, and insignificantly negative for all Nasdaq trades. We conclude that, when sidedness and the number of trades are taken into account, the imbalance variable adds only an inconclusive and essentially meaningless contribution.

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, holding other relevant variables constant, volatility is significantly higher in the first 15 minutes of trading; this would be consistent with opening volatility being a price discovery phenomenon, as others have suggested. The descriptive statistics (Table 2) indicates that, for our sample, HILO is significantly lower in the last half-hour of trading in both exchanges. The regression results are consistent with the descriptive statistics. Holding other relevant variables constant, volatility is significantly lower in the last half-hour of trading.

A further finding of particular interest is that the dummy variable for news days 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 the lagged values of HILO omitted, the news dummy coefficient 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 intervals of high 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, one that conforms to the predictions of models 1 and 2 (but not model 3) that are presented in Table 1. We have previously established that markets are two-sided; we now see that volatility is higher for half-hour windows that fall in the HH cell of the HML matrix. This finding is robust; even with trade imbalance and volume controlled for, volatility is highest in periods when many buyer-initiated and seller-initiated trades cluster together, and when markets are one-sided with many buyer or seller-initiated trades. Volatility is least when markets are two-sided with few buyer-initiated and seller-initiated trades. 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. Due to the higher volatility with two-sided clustering (in the HH cell), we may expect to also find higher trading costs at these times. However, controlling for volatility could change the picture. Further, in intermediated markets, one-sided trade bursts are likely to lead to inventory imbalances for dealers, and to higher trading costs (Chordia et al, 2002). Does sidedness impact trading costs, even after controlling for trade imbalance?

To answer these questions, 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 (instead of PEBAS in the previous regressions) 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 (instead of 3 lags of HILO in the previous regressions) to account for autocorrelation in the bid-ask spread.

We examine descriptive statistics for PEBAS under different conditions of sidedness (i.e. for different cells of the HML matrix) for the NYSE and the Nasdaq stocks. The results are presented in Table 7. For both samples, and for all trade sizes, the mean and median trading costs are generally highest for one-sided markets (the LH, and HL cells). Further, the mean and median trading costs generally increase as we move along the diagonal from LL (few trades) to HH (two-sided, clustered buyer-initiated and seller-initiated trades). In all cases, the bottom two

---

[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).

rows indicate that the mean and median differences in trading costs between the different degrees of sidedness are statistically significant.

The trading cost regression results are given in Table 8 for NYSE and Nasdaq stocks, and for large and all trades. The findings are broadly consistent with the previously discussed descriptive statistics. For both the NYSE and Nasdaq samples, 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) is *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 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.

The effect of sidedness on trading costs, as discussed above, obtains even after controlling for 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 studies that find a similar relation using daily data. For example, Corwin and Lipson (2000) find that the bid-ask spread increases in response to large order imbalances prior to NYSE trading halts, and Chordia et al (2002) find that market liquidity is negatively associated with order imbalances. We also find that trading costs are significantly and negatively related to trading frequency 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. Consistent with prior research, 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 trade imbalance and volume, volatility, news, time-of-day effects, and other microstructure effects accounted for, trade clustering and the sidedness of markets are significant determinants of trading costs. In particular, we find that trading costs are lower when markets are two-sided compared to one-sided markets. These findings conform to the predictions of models 2 and 3 (but not model 1) that are presented in Table 1. Viewed together, the volatility and trading costs results are consistent with model 2, but not with models 1 and 3. This underscores the important impact that investors' having different interpretations of public news and/or different private information signals has on trading and price formation.

6. Additional Investigations

So far, we have shown that, for different exchanges, trade size, time of day, and information conditions, we have shown that the buyer-initiated and seller-initiated trade arrivals are positively correlated, indicating that markets are typically two-sided. In this section, we examine whether markets for smaller stocks are more one-sided than for larger stocks due, for example, to a lack of sufficient investor interest. We further address the concern that our results are peculiar to the post-decimalization period when trade sizes decreased substantially and large orders were more likely to have been broken up, by repeating the analysis for the pre-decimalization period of June 2000. Next, we examine whether trade clustering may be due to the presence of stale limit orders.[24] Finally, we consider alternative definitions of volatility to mitigate the concern that the HILO measure may be contaminated by the bid-ask bounce.

A. Clustering and sidedness for small and large stocks

We split the sample of 41 stocks into 21 small and 20 large stocks, according to their market capitalization as of January 3, 2003. For the small and large stock samples, we separately estimate the 3x3 high-medium-low (HML) matrix of buyer and seller-initiated trade arrivals, using the methodology described in section 3A. The results are in Table 9 for all trades of NYSE and Nasdaq stocks. Considering the sample of all half-hours, there is evidence that the observed and unexpected percent of half-hours in the diagonal cells is higher for the large stocks; this is consistent with intuition that the markets for the small stocks are relatively more one-

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

sided. The first row of the second panel of Table 9 shows that the difference is statistically significant, but it also shows that the difference is only 1.60 percent for NYSE stocks and about 5 percent for Nasdaq stocks. Moreover, for the first 15 minutes of news days, there is no statistically significant difference in the share of half-hours on the diagonal cells for large and small stocks. We conclude that the markets for large stocks are only moderately more two-sided.

Since there is some difference in sidedness of markets for small and large stocks, we next investigate whether this difference is reflected in the way sidedness impacts volatility and trading costs. To this end, we re-estimate our regressions after interacting the sidedness dummy variables with the variable SMALL, which is equal to 1 for the smallest 21 stocks and 0 otherwise. Since aggregate trading activity and trade imbalance may also have a greater impact on small stocks, we further interact SMALL with the number of trades and IMBALANCE variables. The results are in Table 10 for all trades of NYSE and Nasdaq stocks; the dependent variable is HILO in Panel A and PEBAS in Panel B. For brevity, only results for the sidedness dummy variables, trading frequency and IMBALANCE variables are reported.

From Panel A of Table 10, we find that, relative to large stocks, small stock volatility is lower in extreme one-sided markets (i.e. the HL and LH cells) and higher in the extreme two-sided markets (i.e. the HH cells). Recall that, for all stocks, we earlier found that volatility is highest in the HH cells and relatively small in the HL and LH cells. Thus, for small stocks, this tendency is exaggerated. We also find that the impact of trading frequency on volatility is greater for small stocks. Turning to Panel B of Table 10, we find, for small NYSE stocks, that there is little additional impact of sidedness on trading costs. For Nasdaq stocks, however, we find that trading costs are higher for small stocks for half-hours in the HH, HM, MH, HL or LH cells. In other words, whenever there is high trade arrival on one side of the market, trading costs increase for small Nasdaq stocks. The increase in trading costs is highest for small stocks when there is high trade arrival on both sides of the market (i.e. the HH cells).[25] Even considering these additional effects for small stocks, however, it remains true that trading costs are lower in two-sided markets (represented by the diagonal cells) compared to one-sided markets (represented by the HL and LH cells). Thus, while our results show some quantitative

[25] These differences in results for NYSE and Nasdaq stocks may be due to the participation of the NYSE specialist in small stock trades.

difference between large and small stocks in the relationship of sidedness, volatility and trading costs, the qualitative conclusions remain unchanged.

B. Clustering and sidedness in the pre-decimalization period

Trade sizes have decreased substantially after decimalization in January 2001. 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. This suggests that large orders are more likely to have been broken up following decimalization. If large orders are more likely to be one side of the market, markets for large trades may have been more one-sided in the pre-decimalization period. To examine this issue, we repeat our analysis for June 2000 for our sample of 82 stocks. Results using large trades are reported in Table 11. Results using all trades are similar, and are available from the authors.

Panel A of Table 11 shows the distribution of buyer-initiated and seller-initiated trades. As previously, we find that positive unexpected arrivals in the diagonal cells and negative unexpected arrivals in the HL and LH cells; further, the chi-square contribution of diagonal cells is close to 80 percent. These results hold for both NYSE and Nasdaq stocks. Thus, the prevalence of two-sided markets does not appear to be an artifact of the post-decimalization era. Panels B and C of Table 11 show results for regressing HILO and PEBAS on dummy variables for sidedness and clustering. Again, results are similar to those for the post-decimalization era. Volatility is highest in two-sided markets with high arrivals (i.e. the HH cells) whereas trading costs are highest in one-sided markets (i.e. the HL and LH cells). Results for the control variables, not shown to preserver space, are also similar to those reported earlier. Overall, our results appear to be robust to the reduction of trade sizes following decimalization.

C. 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. 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 during a *half-hour*. Thus, for stale limit orders to cause two-sided markets, it must be the case that good news typically follow bad news, or vice versa, within a half-hour. While such a scenario seems unlikely to occur on a regular basis, we nevertheless address the possibility by dividing stocks into two groups according to the average number of trades in a half-hour. Since stale limit orders is more likely to cause trade clustering in fast-moving active markets, the incidence of two-sided markets should be greater for the group of stocks with the highest trading frequency in a half-hour. Our results (which are not reported, but are available upon request) show, however, that the observed and unexpected percent of half-hours along the diagonal cells is similar for the two groups of stocks. Thus, it seems unlikely that our results are primarily due to the presence of stale limit orders.

D. Alternative definitions of volatility

We define volatility as the log ratio of the maximum to the minimum price in a half-hour interval. If the high price is (say) at the ask, and the low price is (say) at the bid, then our measure of volatility may include the bid-ask bounce. To address this concern, we repeat our analysis using, as our definition of volatility, the log ratio of the maximum to the minimum mid-point of the bid-ask quote. The results (not shown but available from the authors) are similar to those shown earlier. In particular, the regression results show that 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.

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 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), the buyer-initiated and seller-initiated trade arrivals are positively correlated, which means that the trade bursts are two-sided. Particularly striking is the extent to which two-sidedness continues to prevail on days characterized by significant 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 on the stock specific level as well.

After having established the prevalence of two-sided trade clustering, we considered the impact of these bursts on price volatility and trading costs. With market conditions, share price, and market cap controlled for, we observe for the array of market conditions considered, that trade bursts lead to higher volatility and to higher trading costs, and that the trading costs are generally higher when markets are one-sided than when they are two-sided.

Evidence of inter-temporal clustering for aggregate trades is seen in the time duration analysis of Hasbrouck (1996) and Dufour and Engle (2000). 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. Rather, we suggest that the explanation lies in 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 yields several implications for trader behavior and market structure.

First, the prevalence of two-sidedness for all trades in general and for large trades in particular, underscores the importance of the liquidity and divergent expectations motives for trading. Previous empirical research has focused on divergent expectations as a motive for trading treasury securities.[26] 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. In light of this, a further investigation into the occurrences of trade clusters and the magnitude of latent liquidity would be desirable.

---

[26] See Fleming and Remolena (1999).

[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.

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

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

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

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

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

15. 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.

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

17. 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.

18. 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.

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

20. Hasbrouck, Joel, 1991a, Measuring the Information Content of Stock Trades, Journal of Finance, 46, 179-207.

21. Hasbrouck, Joel, 1991b, The Summary of Informativeness of Stock Trades: An Econometric Analysis; Review of Financial Studies; 4, 517-595.

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

23. Hendershott, Terrence and Jones, Charles M., 2005, Island Goes Dark: Transparency, Fragmentation, and Regulation, Review of Financial Studies, 18, 743-793.

24. Holthausen and Verrecchia, 1988, The Effect of Sequential Information Releases on the Variance of Price Changes in an Intertemporal Multiasset Market, Journal of Accounting Research, 26, 82-106.

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

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

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

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

29. Lee, C. M.C. and B. Radhakrishna, 2000, Inferring Investor Behavior: Evidence from TORQ data, Journal of Financial Markets, 3, 83-111.

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

31. 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.

32. 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.

33. 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.

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

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

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

37. 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. $\varepsilon_j=P_j n$, 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 • ?
3. Investors trade to rebalance their portfolios (Wang , 1994; He and Wang, 1995; Llorente, Michaely, Saar and Wang, 2002)	• Two-sided markets • If trades cluster, then it occurs on both sides of the market • No relation between sidedness and volatility • No relation between sidedness and trading costs	• Yes • Yes • 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 of 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 a half-hour interval. The remaining measures are computed separately for the sample of all trades (*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. VOLUME and TRADES are total volume and number of trades per half-hour interval. BUYS and SELLS are the number of buy-triggered and sell-triggered trades per half-hour 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 a half-hour 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 first half-hour and the middle half-hours, or between the last half-hour 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 large seller initiated trades in a half-hour interval are statistically independent. The statistical measures are computed separately for the sample of all trades (*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. 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 large buy and large sell triggered trades (e.g. two large buy trades and one large sell trade) was observed, and record them in a buy-sell (BSELL) matrix. To test the independence hypothesis, we use the Pearson chi-square statistic Q_P, which is computed as:

$Q_P = \sum_i \sum_j \frac{(n_{ij} - \varepsilon_{ij})^2}{\varepsilon_{ij}}$, where n_{ij} is the observed frequency of buy-triggered and sell-triggered trade arrivals in row i and column j, the expected frequency 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 trades					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 trades	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 trades					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 trades	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 trades (*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. 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 return from the previous day's closing price to the current day's opening price.

The details of the calculations are as follows. We first count the number of half-hour windows for which each combination of large buy and large sell triggered trades (e.g. two large buy trades and one large 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. 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 $u_{ij} = o_{ij} - e_{ij}$. Finally, 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 $\chi_{ij} = \frac{Q_{ij}}{Q_P}$.

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 large 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 large 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+$\sqrt{\lambda_b}$)

An identical procedure is carried out for large 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 large sell trades for the stock.

The table reports 3 numbers for each cell of the HML matrix: the observed and **unexpected (in bold)** percent of trade arrivals, and the contribution of each cell to Q_P. To obtain these numbers, 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 HL cell, we sum o_{ij}, u_{ij} and χ_{ij} over all cells (i,j) of the BSELL matrix that are mapped into the HL 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 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 B 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: Distribution of buyer and seller-initiated trades

Panel A. NYSE stocks

Observed and unexpected (in bold) percent of half-hours, and chi-square share (in %, last entry) of cells in High-Medium-Low (HML) buy-sell trade arrival matrix

	All trades					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	6.19	6.16	16.42	25.51	8.84	2.85	2.80	10.96	24.20	17.25
	-6.69	**-6.80**	**7.11**	**7.47**	**1.09**	**-5.81**	**-5.80**	**6.38**	**7.52**	**2.29**
	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	8.45	8.71	13.85	21.55	8.34	6.96	6.84	9.76	20.44	12.04
	-4.35	**-3.98**	**4.18**	**4.44**	**0.30**	**-2.91**	**-3.37**	**4.07**	**2.08**	**-0.14**
	8.73	9.57	36.83	16.42	2.98	3.65	3.45	69.47	3.90	2.78
Last 15 minutes, all days	10.58	11.13	13.83	22.29	6.81	6.94	6.65	11.22	20.06	11.47
	-3.13	**-2.54**	**2.95**	**4.57**	**-0.43**	**-4.32**	**-3.99**	**4.25**	**1.94**	**0.15**
	11.32	13.72	29.40	19.74	4.38	4.86	6.66	59.02	7.88	5.82
First 15 minutes, news days	7.89	9.33	12.48	21.52	8.11	6.24	6.31	10.19	21.95	11.25
	-3.95	**-3.76**	**3.28**	**4.29**	**-0.14**	**-3.51**	**-4.14**	**4.73**	**2.38**	**-0.55**
	9.20	8.56	22.47	18.59	5.71	7.46	5.69	47.81	6.06	4.58

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

	All trades			Large trades		
Null hypothesis	**Estimate**	**T-statistic**	**P-value**	**Estimate**	**T-statistic**	**P-value**
μ(LL+MM+HH)=0, All half-hours – First 15 minutes	7.04**	6.66	0.0000	10.18**	9.67	0.0000
μ(LH+HL)=0, All half-hours – First 15 minutes	-4.82**	-7.35	0.0000	-8.16**	-14.04	0.0000
μ(LL+MM+HH)=0, All half-hours – Last 15 minutes	9.98**	9.71	0.0000	7.18**	6.64	0.0000
μ(LH+HL)=0, All half-hours – Last 15 minutes	-9.36**	-12.81	0.0000	-7.95**	-13.67	0.0000

Mean difference in observed percent of half-hours, for first 15 minutes of news dyas versus first 15 minutes of non-news days

	All trades			Large trades		
Null hypothesis	**Estimate**	**T-statistic**	**P-value**	**Estimate**	**T-statistic**	**P-value**
μ(LL+MM+HH)=0, First 15 minutes, all days – First 15 minutes, news days	1.69	0.84	0.4034	-1.19	-0.58	0.5617
μ(LH+HL)=0, First 15 minutes, all days – First 15 minutes, news days	-0.04	-0.03	0.9730	1.24	1.12	0.2650

Table 4 (continued)

Panel B. Nasdaq stocks

Observed and unexpected (in bold) percent of half-hours, and chi-square share (in %, last entry) of cells in High-Medium-Low (HML) buy-sell trade arrival matrix

	All trades					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	7.51	7.51	19.99	36.75	3.44	3.02	2.79	13.33	35.07	11.33
	-9.16	**-9.45**	**9.35**	**9.92**	**0.66**	**-8.08**	**-8.10**	**8.21**	**10.63**	**2.65**
	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	9.80	9.73	18.83	33.39	3.01	4.24	4.55	14.35	42.41	6.80
	-7.29	**-7.08**	**7.18**	**7.66**	**0.48**	**-8.82**	**-8.55**	**8.68**	**10.48**	**1.87**
	10.49	9.07	33.50	23.91	2.03	3.40	2.69	69.33	8.81	2.37
Last 15 minutes, all days	12.79	13.78	18.02	28.18	2.18	6.01	5.75	14.35	32.82	7.20
	-4.51	**-4.64**	**4.13**	**4.99**	**-0.03**	**-7.30**	**-7.33**	**6.88**	**8.97**	**1.22**
	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	9.24	9.77	20.02	37.19	1.48	3.00	3.67	16.01	50.30	4.59
	-8.05	**-8.74**	**7.99**	**8.46**	**-0.36**	**-10.60**	**-10.51**	**10.18**	**12.63**	**1.76**
	9.88	10.47	24.46	32.46	1.51	3.58	4.60	46.58	19.63	4.74

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

	All trades			Large trades		
Null hypothesis	**Estimate**	**T-statistic**	**P-value**	**Estimate**	**T-statistic**	**P-value**
μ(LL+MM+HH)=0, All half-hours – First 15 minutes	4.96**	4.16	0.0000	-3.83**	-3.02	0.0026
μ(LH+HL)=0, All half-hours – First 15 minutes	-4.51**	-6.45	0.0000	-2.98**	-6.33	0.0000
μ(LL+MM+HH)=0, All half-hours – Last 15 minutes	11.81**	10.53	0.0000	5.36**	4.55	0.0000
μ(LH+HL)=0, All half-hours – Last 15 minutes	-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 news days versus first 15 minutes of non-news days

	All trades			Large trades		
Null hypothesis	**Estimate**	**T-statistic**	**P-value**	**Estimate**	**T-statistic**	**P-value**
μ(LL+MM+HH)=0, First 15 minutes, all days – First 15 minutes, news days	-3.21	-1.36	0.1739	-7.21**	-2.79	0.0054
μ(LH+HL)=0, First 15 minutes, all days – First 15 minutes, news days	0.41	0.30	0.7661	2.12*	2.54	0.0112

Table 5: Effect of Sidedness and Clustering on Volatility

The table shows the distribution of volatility of large trades for 41 NYSE and 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. Statistics are reported separately for the sample of all trades (*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 volatility is HILO, which is the log of the highest to the lowest price in a half-hour interval. The column headings HH, LH, HL, MM, LL refer to cells in the 3x3 High-Medium-Low (HML) buy-sell matrix (e.g., HH refers to the HIGH BUY, HIGH SELL cell). Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. Starred mean and median values are significantly different from zero. *Tests of hypotheses* refer to tests of whether the mean (median) in column *j* is significantly different from the mean (median) in column LL, where *j*= MM, (LH, HL), HH. ** indicates significance at the 1 per cent level or less; * indicates significance at the 5 percent level or less.

Panel A. NYSE stocks

	All trades				Large trades			
	LL	**MM**	**LH, HL**	**HH**	**LL**	**MM**	**LH, HL**	**HH**
No. of obs.	13,334	4,854	6,780	9,016	12,614	9,470	3,099	6,015
Mean	0.52**	0.73**	0.75**	1.23**	0.59**	0.75**	0.78**	1.22**
Std. dev.	0.37	0.51	0.55	0.99	0.42	0.54	0.63	1.10
Max	3.92	5.09	5.16	36.45	4.56	5.09	7.93	36.45
Median	0.44**	0.61**	0.62**	1.01**	0.50**	0.63**	0.62**	0.96**
Min	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.04
Tests of Hypotheses								
Difference in mean, from LL		0.21**	0.23**	0.71**		0.16**	0.19**	0.63**
Difference in median, from LL		0.17**	0.18**	0.57**		0.13**	0.12**	0.46**

Panel B. Nasdaq stocks

	All trades				Large trades			
	LL	**MM**	**LH, HL**	**HH**	**LL**	**MM**	**LH, HL**	**HH**
No. of obs.	19,693	1,889	8,249	10,974	18,768	3,188	6,220	7,317
Mean	0.60**	0.82**	0.86**	1.56**	0.69**	0.82**	0.89**	1.54**
Std. dev.	0.39	0.55	0.56	1.02	0.48	0.65	0.63	1.09
Max	4.33	5.57	5.46	14.19	6.65	8.25	5.57	14.19
Median	0.53**	0.70**	0.75**	1.32**	0.59**	0.66**	0.75**	1.26**
Min	0.00	0.03	0.00	0.04	0.00	0.00	0.00	0.00
Tests of Hypotheses								
Difference in mean, from LL		0.22**	0.26**	0.96**		0.13**	0.20**	0.85**
Difference in median, from LL		0.17**	0.22**	0.79**		0.07**	0.16**	0.67**

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 trades (*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 a half-hour interval
- IMBALANCE: the absolute value of the number of buyer-initiated trades minus the number of seller-initiated trades, as a proportion of the total number of trades of that type (all or large)
- 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. 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 trades

	All trades		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 trades

	All trades		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: Effect of Sidedness and Clustering on Trading Costs

The table shows the distribution of trading costs for 41 NYSE (Panel A) and Nasdaq (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 trades (*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 spread in a half-hour. PEBAS is Q*(P- M)/M, where P is the trade price, and Q is +1 (-1) for a buyer (seller) initiated trade and M is the quote mid-point. The column headings HH, MH, LH, HL, MM, LL refer to cells in the 3x3 High-Medium-Low (HML) matrix (e.g., HH refers to the HIGH BUY, HIGH SELL cell). Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. Starred mean and median values are significantly different from zero. *Tests of hypotheses* refer to tests of whether the mean (median) in column *j* is significantly different from the mean (median) in column LL, where *j*= MM, (LH, HL), HH. ** indicates significance at the 1 per cent level or less; * indicates significance at the 5 percent level or less.

Panel A. NYSE stocks

	All trades				Large trades			
	LL	**MM**	**LH, HL**	**HH**	**LL**	**MM**	**LH, HL**	**HH**
No. of obs.	13,334	4,854	6,780	9,016	7,560	9,470	3,099	6,015
Mean	0.080**	0.087**	0.094**	0.100**	0.104**	0.116**	0.135**	0.121**
Std. dev.	0.081	0.082	0.094	0.089	0.144	0.167	0.190	0.117
Max	1.392	1.200	1.370	0.950	1.506	3.384	3.626	1.454
Median	0.056**	0.063**	0.067**	0.072**	0.059**	0.070**	0.076**	0.083**
Min	0.000	0.008	0.007	0.009	0.000	0.000	0.000	0.005
Tests of Hypotheses								
Difference in mean, from LL		0.007**	0.014**	0.020**		0.012**	0.031**	0.017**
Difference in median, from LL		0.007**	0.011**	0.016**		0.011**	0.017**	0.024**

Panel B. Nasdaq stocks

	All trades				Large trades			
	LL	**MM**	**LH, HL**	**HH**	**LL**	**MM**	**LH, HL**	**HH**
No. of obs.	13,334	4,854	6,780	9,016	7,560	9,470	3,099	6,015
Mean	0.063**	0.070**	0.065**	0.064**	0.065**	0.073**	0.079**	0.067**
Std. dev.	0.051	0.049	0.050	0.043	0.060	0.055	0.064	0.044
Max	0.556	0.313	0.416	0.418	0.767	0.602	0.588	0.442
Median	0.045**	0.053**	0.048**	0.051**	0.046**	0.056**	0.058**	0.054**
Min	0.000	0.012	0.006	0.011	0.000	0.006	0.000	0.012
Tests of Hypotheses								
Difference in mean, from LL		0.007**	0.002**	0.001		0.007**	0.014**	0.001
Difference in median, from LL		0.008**	0.003**	0.006**		0.009**	0.012**	0.008**

Table 8: 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 trades (*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: the absolute value of the number of buyer-initiated trades minus the number of seller-initiated trades, as a proportion of the total number of trades of that type (all or large)
- 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 8 (continued)

Panel A: PEBAS, NYSE stocks

	All trades		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 trades		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 9: Distribution of Buyer and Seller-initiated Trades for Large and Small Stocks

Each cell of Panel A 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. L, M and H stand for low, medium and high trade arrivals, with the first letter of a pair representing buyer-initiated trades. For example, HL stands for high buyer-initiated and low seller-initiated trade arrivals. Results are shown for all days, and for the first 15 minutes on days with news. News days for a stock are the 30 percentile of days with the largest values of ACLOP, the absolute value of the log excess return from the previous day's closing price to the today's opening price. Results are computed using *all* trades, but separately for the sample of the 21 smallest stocks and a sample of the 20 largest stocks by market value. Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. In Panel B, we show *t*-statistics and *p*-values for the null hypotheses that the difference in mean (μ) of observed percent of half-hours between small and large stocks is zero for (1) the diagonal cells and (2) the HL and LH cells. ** (*) 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 according to their closing price and market value on December 31 2002. Details of the calculations are as described in previous tables

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

	All trades, 21 smallest stocks					All trades, 20 largest stocks				
	H,L	L,H	H,H	L,L	M,M	H,L	L,H	H,H	L,L	M,M
All half-hours NYSE stocks	6.30	6.36	15.68	25.31	9.31	6.02	6.01	17.21	25.72	8.35
	-6.42	**-6.35**	**6.73**	**7.15**	**1.12**	**-7.19**	**-7.05**	**7.51**	**7.79**	**1.06**
	3.70	3.04	76.23	6.53	1.01	2.75	2.61	72.90	10.88	1.04
All half-hours Nasdaq stocks	7.50	7.35	17.66	35.97	4.18	7.53	7.68	22.44	37.57	2.67
	-8.51	**-8.25**	**8.33**	**9.10**	**0.68**	**-10.45**	**-10.12**	**10.43**	**10.79**	**0.65**
	2.49	2.50	82.29	5.19	0.53	3.74	4.24	67.77	14.18	0.70
First 15 minutes, news days, NYSE	7.56	9.10	11.06	23.25	8.54	8.68	9.26	14.56	22.79	5.44
	-3.81	**-3.98**	**2.76**	**5.10**	**0.07**	**-4.84**	**-4.44**	**4.03**	**4.42**	**-0.83**
	9.58	8.32	23.81	18.20	5.58	9.23	9.08	22.06	21.05	3.66
First 15 minutes, news days, Nasdaq	8.79	8.57	18.79	40.96	1.10	8.09	9.71	20.15	32.50	2.79
	-8.67	**-9.04**	**8.43**	**8.66**	**-0.68**	**-8.24**	**-7.96**	**7.77**	**8.63**	**0.20**
	9.52	10.51	25.00	33.32	0.91	8.17	11.50	24.18	30.22	2.29

Panel B: Mean difference in observed percent of half-hours, between large and small stocks

	All trades, NYSE			All trades, Nasdaq		
Null hypothesis	**Estimate**	**T-statistic**	**P-value**	**Estimate**	**T-statistic**	**P-value**
μ(LL+MM+HH)=0, All half-hours, 20 largest – 21 smallest stocks	0.97	1.60	0.1106	4.88**	7.37	0.0000
μ(LH+HL)=0, 20 largest – 21 smallest stocks	-0.63*	-2.10	0.0363	0.37	1.13	0.2596
μ(LL+MM+HH)=0, First 15 minutes, news days, 20 largest – 21 smallest stocks	0.04	0.01	0.9910	-4.93	-1.20	0.2291
μ(LH+HL)=0, First 15 minutes, news days, 20 largest – 21 smallest stocks	1.30	0.58	0.5628	0.22	0.10	0.9216

Table 10: Trading Cost and Volatility Regressions for Large and Small Stocks

The table shows results from a regression of volatility and trading costs on dummy variables for sidedness and clustering. The sample is all trades for 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. In Panel A, the dependent variable is HILO, equal to the log of the highest to the lowest price in a half-hour interval. In Panel B, the dependent variable is PEBAS, the average proportional effective bid-ask half-spread in a half-hour interval, equal to Q*(P- M)/M, where P is the trade price, and Q is +1 (-1) for a buyer (seller) initiated trade. The explanatory variables are dummy variables that refer to cells in the 3x3 High-Medium-Low (HML) buy-sell matrix (e.g., HH refers to the HIGH BUY, HIGH SELL cell):

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 LM and ML. The regression also includes the following control variables:

- Log of the number of trades in a half-hour interval
- IMBALANCE: the absolute value of the number of buyer-initiated trades minus the number of seller-initiated trades, as a proportion of the total number of trades of that type (all or large)

The above variables are interacted with SMALL, which is equal to 1 for the 21 smallest stocks and is zero otherwise.

The following control variables (which do not interact with SMALL) are included in the regression but results are not reported to save space:

- Log of the previous day's closing price
- 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.
- HILO (when the dependent variable is PEBAS) or PEBAS (when the dependent variable is HILO)
- 3 lags of HILO (when the dependent variable is HILO) or PEBAS (when the dependent variable is 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 10 (continued)

Panel A: Dependent variable is HILO

	All trades, NYSE		All trades, Nasdaq	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Intercept	-0.0390	-0.98	-0.7590**	-22.57
Dummy1 (LL)	-0.0510**	-10.20	-0.0450**	-7.40
Dummy2 (MM)	0.0544**	6.31	0.0487**	3.74
Dummy3 (HL,LH)	0.0490**	6.45	0.0637**	8.10
Dummy4 (MH,HM)	0.1170**	14.97	0.1260**	13.42
Dummy5 (HH)	0.2208**	21.42	0.2910**	29.18
Log of Number of trades	0.1484**	42.15	0.2083**	58.79
IMBALANCE	-0.0020	-1.14	0.0013	0.55
Dummy1 (LL)*SMALL	-0.0150*	-1.98	-0.0250**	-2.87
Dummy2 (MM) *SMALL	0.0019	0.15	-0.0230	-1.23
Dummy3 (HL,LH) *SMALL	-0.0220	-1.92	-0.0840**	-6.97
Dummy4 (MH,HM) *SMALL	0.0055	0.47	-0.0240	-1.60
Dummy5 (HH) *SMALL	0.0344*	2.10	0.0556**	3.35
Log Number of trades*SMALL	0.0084**	3.63	0.0275**	11.47
IMBALANCE*SMALL	-0.0040	-1.35	-0.0070*	-2.02
Adjusted R-squared	0.49		0.56	
Number of observations	61,899		62,069	

Panel B: Dependent variable is PEBAS

	All trades, NYSE		All trades, Nasdaq	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Intercept	0.1055**	16.93	0.0767**	29.40
Dummy1 (LL)	-0.0010	-1.72	-0.0010**	-3.32
Dummy2 (MM)	-0.0020*	-2.41	-0.0010*	-1.97
Dummy3 (HL,LH)	0.0025**	2.92	-0.0004	-1.30
Dummy4 (MH,HM)	-0.0010	-1.49	-0.0020**	-4.78
Dummy5 (HH)	-0.0030**	-3.59	-0.0060**	-14.15
Log of Number of trades	-0.0120**	-18.57	-0.0080**	-26.58
IMBALANCE	0.0007**	3.22	0.0003**	2.92
Dummy1 (LL)*SMALL	-0.0020	-1.34	0.0006	1.04
Dummy2 (MM) *SMALL	0.0003	0.13	0.0018	1.38
Dummy3 (HL,LH) *SMALL	0.0031	1.49	0.0039**	4.96
Dummy4 (MH,HM) *SMALL	0.0029	1.59	0.0031**	3.57
Dummy5 (HH) *SMALL	0.0039*	2.15	0.0056**	7.30
Log Number of trades*SMALL	-0.0005	-1.59	-0.0005**	-4.08
IMBALANCE*SMALL	-0.0002	-0.33	0.0012**	5.83
Adjusted R-squared	0.53		0.77	
Number of observations	61,896		62,067	

Table 11: Distribution of Buyer and Seller-initiated Trades and Regression Analysis, for the Pre-Decimalization Period

Each cell of Panel A 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. L, M and H stand for low, medium and high trade arrivals, with the first letter of a pair representing buyer-initiated trades. For example, HL stands for high buyer-initiated and low seller-initiated trade arrivals. Statistics are shown for all days, and for the first 15 minutes on days with news. Statistics are shown for the entire sample and for the first 15 minutes of days with news. News days for a stock are the 30 percentile of days with the largest values of ACLOP, the absolute value of the log excess return from the previous day's closing price to the current day's opening price. Results are computed using only *large* trades, defined as those in the top 10 percentile of the dollar value of trades of a stock in the sample. Buyer and seller initiated trades are determined using the Lee-Ready (1991) algorithm. The details of the calculations are as described in previous tables

Panel B shows results from regressions on dummy variables for sidedness and clustering. In Panel B, the dependent variable is HILO, equal to the log of the highest to the lowest price in a half-hour interval. In Panel C, the dependent variable is PEBAS, the average proportional effective bid-ask half-spread in a half-hour interval, equal to Q*(P- M)/M, where P is the trade price, and Q is +1 (-1) for a buyer (seller) initiated trade. The explanatory variables are dummy variables that refer to cells in the 3x3 High-Medium-Low (HML) buy-sell matrix (e.g., HH refers to the HIGH BUY, HIGH SELL cell):

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 LM and ML. The regression also includes the following control variables:

- Log of the number of trades in a half-hour interval
- IMBALANCE: the absolute value of the number of buyer-initiated trades minus the number of seller-initiated trades, as a proportion of the total number of trades of that type (all or large)

A number of other control variables are included in the regression but results are not reported to save space. These are: Log of the previous day's closing price; 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; HILO (when the dependent variable is PEBAS) or PEBAS (when the dependent variable is HILO); 3 lags of HILO (when the dependent variable is HILO) or PEBAS (when the dependent variable is 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. The sample is all trades for 41 NYSE and 41 Nasdaq stocks during January 2 to May 31 2003.

Table 11 (continued)

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

	Large trades, NYSE stocks					Large trades, Nasdaq stocks				
	H,L	**L,H**	**H,H**	**L,L**	**M,M**	**H,L**	**L,H**	**H,H**	**L,L**	**M,M**
All half-hours	6.23	6.26	6.80	35.35	17.26	3.93	3.67	11.66	39.28	12.44
	-3.31	**-3.20**	**3.34**	**4.22**	**1.48**	**-6.83**	**-6.85**	**6.94**	**9.31**	**2.84**
	4.74	4.67	74.73	1.95	1.87	2.50	2.69	79.44	4.23	1.94
First 15 minutes, news days	10.14	9.96	10.64	35.83	9.27	8.99	4.62	18.91	47.07	10.32
	-2.49	**-3.78**	**4.00**	**1.38**	**-2.08**	**-7.52**	**-9.96**	**9.78**	**8.76**	**3.74**
	16.69	10.96	29.11	10.51	4.52	10.05	9.52	31.67	20.54	9.54

Panel B: Dependent variable is HILO

	Large trades, NYSE		Large trades, Nasdaq	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Intercept	0.2696**	3.40	-0.7270**	-7.19
Dummy1 (LL)	-0.1230**	-4.31	-0.1850**	-7.36
Dummy2 (MM)	-0.0030	-0.12	0.0763*	2.53
Dummy3 (HL,LH)	0.1512**	7.26	0.0438	1.31
Dummy4 (MH,HM)	0.2268**	8.70	0.1918**	6.24
Dummy5 (HH)	0.4139**	9.27	0.4864**	9.57
Log of Number of trades	0.2911**	12.42	0.2875**	16.00
IMBALANCE	-0.0030	-0.70	0.0080	1.30
Adjusted R-squared	0.39		0.40	
Number of observations	8,093		8,979	

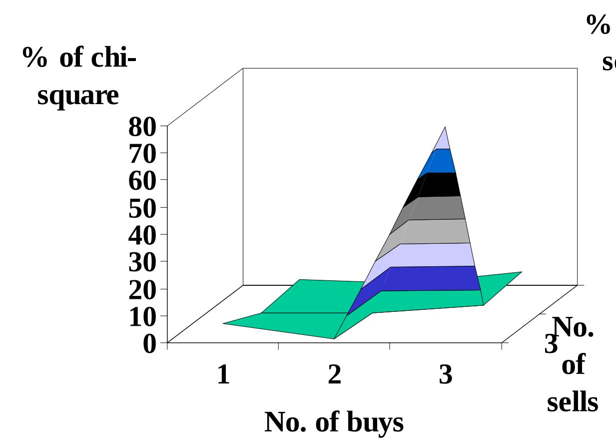
Panel C: Dependent variable is PEBAS

	Large trades, NYSE		Large trades, Nasdaq	
Explanatory variable	**Estimate**	**t-statistics**	**Estimate**	**t-statistics**
Intercept	0.5824**	11.46	0.3163**	17.66
Dummy1 (LL)	0.0052	0.31	0.0075	1.89
Dummy2 (MM)	-0.0340*	-1.98	-0.0080*	-2.02
Dummy3 (HL,LH)	0.0395**	2.65	0.0072	1.00
Dummy4 (MH,HM)	-0.0009	-0.06	-0.0060	-1.35
Dummy5 (HH)	-0.0120	-0.72	-0.0140**	-3.44
Log of Number of trades	-0.0200*	-2.57	-0.0300**	-13.76
IMBALANCE	-0.0050	-1.34	0.0010	1.19
Adjusted R-squared	0.16		0.53	
Number of observations	5,325		6,927	

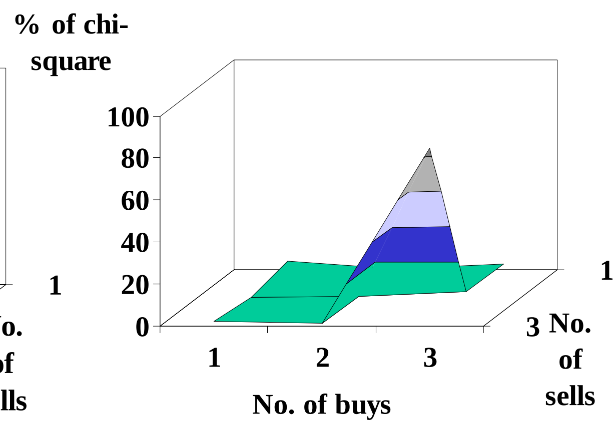
Figure 1: Distribution of buyer and seller-initiated trades

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.

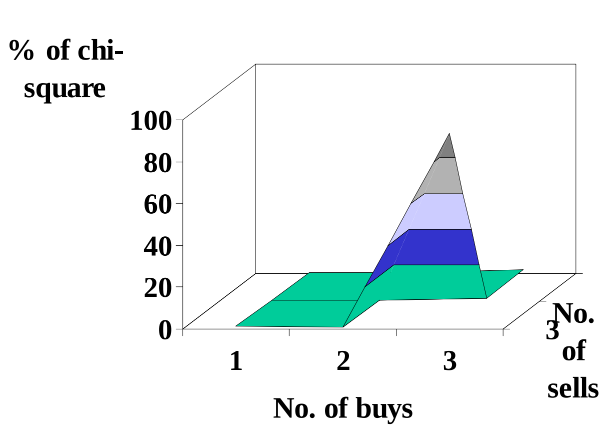
Most Active NYSE stock GM



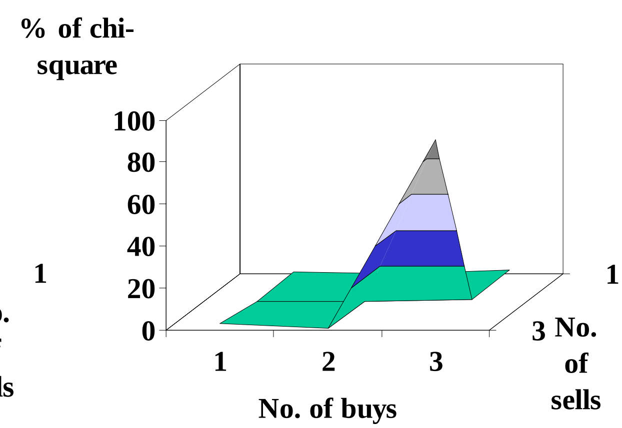
Most Active Nasdaq stock QCOM



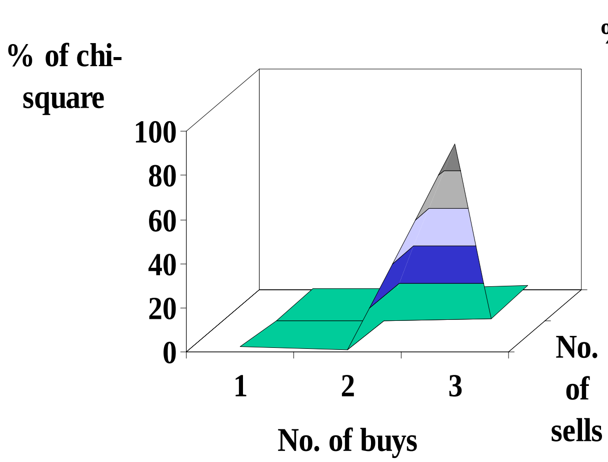
Medium Active NYSE stock CMX



Medium Active Nasdaq stock EXPD



Least Active NYSE stock CUZ



Least Active Nasdaq stock CBCF