The Relationship between the Information Content of Trades and Frequency of Public Information Release: The Role of Informed and Uninformed Trading

Srinivasan Sankaraguruswamy*

Jianfeng Shen

Takeshi Yamada

(National University of Singapore)

Abstract

In this paper, we empirically document that the information content of trades in firms with more frequent public information release is lower on average. We further show that both informed and uninformed traders trade more in firms with more frequent public release. However, the trading by uninformed traders is of a greater order of magnitude than the trading by informed traders. This has the effect of reducing the information content of trades in firms with more frequent public releases. Our findings highlight the important role of public information in leveling the playing field for all investors.

* Corresponding author
NUS School of Business
National University of Singapore
1 Biz Link #2-24 Biz 1
Singapore 119572
bizsrini@nus.edu.sg

Key Word: Public Information Release, Information Content of Trades, PIN, Uninformed Trading

JEL: G14, D82

1. Introduction

In this paper we study whether order flow in firms with more frequent public information release have higher or lower information content. Further, we investigate the role of informed trading and uninformed trading on the relation between frequency of public information release and the information content of trades. The information-based paradigm of security price formation proposes that the evolution of prices is a function of the information content of trades. How much information a trade contains is determined by the interaction between differentially-informed traders (see, e.g., Glosten and Milgrom, 1985; Kyle, 1985; and Easley and O'Hara, 1987). When the information difference or information asymmetry between informed and uninformed traders is high, and when the intensity of informed trading is high, then the information content of trades is high. However, greater trading by uninformed investors allows informed investors to better hide their trades, leading to lower information content of trades (Kyle, 1985). Public information release could increase or decrease trading by both informed and uninformed traders, so, it is not clear what the net effect would be on the information content of trades.[1]

Informed traders may gather private information *in anticipation* of the public information release and take positions to profit from this information (Bagnoli and Watts, 1998). Further, release of public information also affords informed traders the opportunity to gather private information *after* the release of the public information (Kim and Verrecchia, 1994). Both the above strategies increase the proportion of trades by informed traders. However, if informed traders deem public information to be a substitute for private information generation then they are less likely to gather private information and trade on it prior to a public information release (Grundy and McNichols, 1989). Hence, public information release has the effect of either increasing or decreasing the proportion of trading by informed traders.[2]

Public information release could also affect the incentive of uninformed traders to trade. Uninformed traders, more likely to be small and individual investors, are resource constrained and cannot gather costly private information. Hence they may rely on public

[1] We discuss the effect of possible endogenous relation between informed trading and information content of trades in a robustness section.
[2] See Livne (2000) for a model that incorporates both these characteristics of public information release.

information release to make their trading decision and their trades may be clustered in firms with more frequent public information release (Lee, 1992; Barber and Odean, 2005). However, if uninformed traders conjecture that informed traders are more likely to trade before or after a public information release then they may stay away from trading in such stocks to avoid trading against better informed traders. In sum, whether uninformed traders prefer to or avoid trading in firms with frequent public information release is an empirical question.

We test whether frequency of information release is related to the information content of trades, and what role informed and uninformed trading plays in mediating the above relationship in three steps. The first step involves establishing a relation between the frequency of public information release and the information content of trades. In the second step we seek to relate the frequency of public information release with the intensity of informed and uninformed trading. This would clarify the role of public information release in attracting informed and uninformed traders. Finally, we show that the relation between frequency of public information release and information content of trades is mediated by the intensity of informed and uninformed trading.

We test the above conjectures on a sample of 1028 firms whose common stocks are traded on the NYSE during the calendar year 2004. We measure the information content of trades by the Kyle's *Lambda*, which measures the sensitivity of prices to order flow. Our measure of public information release is all firm-specific news appearing on www.MarketWatch.com.[3] This website receives news stories from over 20 sources including, Reuters, and the firms' own press releases. Such a broad coverage of news allows us to make general statements about the effect of public information release on the information content of trades, instead of focusing on a specific news event such as earnings announcement. We measure the intensity of informed trading following the model proposed by Easley et al. (1996). This model describes the probability of an information event and the arrival rates of informed and uniformed trades, and calculates the resultant probability of informed trading.

[3] www.MarketWatch.com is a financial information services provider wholly owned by Dow Jones & Company, Inc. Its news sources include company press releases via Business Wire, PR Newswire, Market Wire, and Prime Zone, SEC filings by Edgar Online, and other sources such as Reuters, New York Times, CBS News, FT.com, TheStreet.com, The Wall Street Journal Online, among others.

We find that the information content of trades is *lower* for firms with *greater* frequency of public information release, measured as the number of days with news on www.MarketWatch.com, in the cross section. Further tests indicate that greater frequency of public information release leads to both higher informed trading and higher uninformed trading. However, the increase in uninformed trading is much higher than that in informed trading in firms with greater frequency of public information release. This difference in the sensitivity between informed and uninformed trading to public information release, accounts for the lower information content of trades in firms with more frequent public information release. Indeed the negative relationship between the frequency of public information release and the information content of trades no longer exists when the relative intensity of informed trading versus uninformed trading is also included in the model.

Our study contributes to the understanding of the effects of public information release on investors' trading behavior and price formation. We document that firms with greater frequency of public information release have, on average, lower information content of trades, although prior empirical research shows that information content of trades increases just before and immediately after a public information event.[4] Information content of trades has been related to the adverse selection costs and the liquidity costs of trading, which in turn affect the cost of capital and expected returns.[5] Our finding that the frequency of public information release decreases the information content of trades highlights the role of public information release in reducing the cost of capital and thus supports the efforts by the SEC to increase public disclosure of information. Further, we show that the increased trading in firms with more frequent public information release is driven by uninformed investors who base their trading

[4] The information content of trades is found to increase before dividend announcements (Koski and Michaley, 2000) and after an earnings announcement (Krinsky and Lee, 1996; Cong, Hoitash and Krishnan, 2005). However, Ronaldo (2004) finds that trading around news releases is characterized by relatively small adverse selection component of bid ask spreads. Graham et al. (2006) document that the information content of trades decreases after unanticipated news announcements, but remains unchanged after anticipated news announcements.

5 Finance and accounting literature has shown that information asymmetry impedes investment and increases cost of capital (see, e.g., Diamond and Verrecchia, 1991; Easley and O'Hara, 2004; among others. See Verrecchia, 2001 for a survey of this literature). Easley, Hvidkjaer and O'Hara (2002, 2005) show that information asymmetry is a factor that impacts prices systematically and Pastor and Stambaugh (2003) find that liquidity impacts prices in the cross section.

decisions on public information rather than on private information. This finding supports the SEC's objective to increase disclosure of public information and hence level the playing field among all investors, especially uninformed investors.

The rest of the paper is organized as follows, section 2 describes the sample and data, section 3 presents the empirical results, and section 4 concludes the paper.

2. Sample Selection

2.1 Sample

Our final sample comprises of 1028 firms traded on the NYSE during the calendar year 2004. We start with a sample of 1535 common stocks traded on NYSE during year 2004. We include only NYSE stocks in our analysis so as to abstract away from differences in market making mechanisms between different exchanges. Stocks other than common stocks, i.e., those with share code other than 10 and 11, are excluded, since their trading characteristics might differ from those of ordinary equities.[6] Consistent with extant literature the following criteria need to be satisfied for stocks to be included in our sample:

1. The stocks were listed on the NYSE during the whole year.
2. The stock's monthly closing price for calendar 2004 should not be lower than $1 or higher than $999 so as to avoid the influence of penny stocks and extremely high-price stocks in which uninformed investors are unlikely to trade.
3. To be included stocks should have at least 60 days of trading during the year and have less than 20 consecutive days with missing trading during regular trading hours. This helps mitigate the impact of thinly traded stocks on the computations relating to the information content of trades.

After applying the above filters, 1411 stocks are left in the sample. We further require the stocks to have news appearing on www.marketwatch.com during 2004, which restricts the sample to 1331 stocks. In calculating the arrival rates of informed and uninformed investors using the Easley et al. (1996) model, 240 stocks with extremely

[6] The stocks that are dropped are in the following categories: certificates, American Depositary Receipts, shares of beneficial interest, units, companies incorporated outside the U.S., Americus Trust components, closed-end funds, preferred stocks, and Real Estate Investment Trusts.

large number of trades per day are excluded due to inability to estimate *PIN* for such stocks. For 63 stocks we are unable to find any analyst forecasts and so we delete them from our sample.[7] We are left with a final sample of 1028 firms.

2.2 Variable definitions

2.2.1 Frequency of public information release

Our source for public news is all the news items appearing on www.MarketWatch.com for each firm. www.MarketWatch.com is a financial information services provider wholly owned by Dow Jones & Company, Inc. There are two merits in using news releases on www.MarketWatch.com as a proxy for firm-specific public information. First, www.MarketWatch.com provides a broad coverage of relevant firm-specific public news, including press releases, SEC filings and other news covered by high-profile media sources such as Reuters, The Wall Street Journal Online, New York Times, CBS News, FT.com, among others. Second, the news on this website is displayed in a manner that allows us to easily search and download the date, time, and headline for each news item using an automated computer algorithm, thus making it easy to gather a large amount of data.

Frequency of public information release is measured as the number of days with news (*Newsdays*) during the sample year 2004 appearing on www.MarketWatch.com for each firm. The choice of number of news days rather than number of news items to measure the frequency of public information release is based on three considerations. First, the use of number of news days gets around the problem of double counting the same news story reported from different news sources which appear as multiple news items on the website. Thus, higher number of news releases during a specific day may not necessarily imply greater flow of information. Second, analyzing a large database of news presents a challenge in terms of assessing the relative importance of news and independence of news items. Since we gather headlines of the news items and do not analyze these for content, we cannot tell, ex ante, whether one news release is more important than another news release. Third, sometimes a news story may describe the previous days' trading and price movement in a stock. This news release does not contain

[7] As a robustness test we include these 63 stocks with zero analyst following and treat them as having one analyst following the stock. Our results remain the same.

any new information and adds noise to our news variable. It may also introduce an endogenous relation between public news and the information content of trades. We check to what extent of this problem exists in our data by aggregating all the news items into categories according to specific key words appearing in the headline of the news release. For example, to delineate news releases that relate to prices and trades of a particular stock and the market as a whole, we search for words like, market, trading, prices, and volume. We find that the proportion of news releases with the above key words in their headlines to the total number of news releases is less than 0.5%. We delete these observations from our sample of news releases. The use of number of news days rather than number of news items mitigates any potential bias introduced by the above considerations. However, this benefit is offset by a loss of information from assigning the same value of 1 to many or one news item.

Table 2 describes the distribution of public news for 1362 NYSE common stocks which have news stories appearing on www.MarketWatch.com during year 2004. The median (mean) value of the number of days with news (*Newsdays*) is 55 (70). The distribution of the number of public news released is skewed and dominated by a few firms which have news almost every day. [8] The median (mean) number of news items per firm appearing in a year is 98 (177). The skewness of the number of news items is almost double that of the skewness of the number of news days. Hence counting news days instead of news items is a better description of the data.

2.2.2 Information content of trades

To estimate the information content of trades, we first classify all valid trades as either buyer-initiated or seller-initiated according to the algorithm proposed by Lee and Ready (1991). The Lee-Ready algorithm deems a trade above (below) the mid point as a buy (sell). A trade at the midpoint of the quoted spread is classified by using either a 'tick' test or a 'zero-tick' test. Specifically, a trade is classified as buy (sell) if it is an up-tick (down-tick), i.e., it is transacted at a higher (lower) price than the previous trade. If there is no price change then the current trade is classified as a buy (sell) if the immediately previous trade is an up-tick (down-tick). Trade and quote data are obtained

[8] Our results hold even if we exclude firms falling in the top and bottom 1% of the distribution of the frequency of public news days or number of news items.

from the NYSE TAQ database. We include all trades except for corrected trades, trades settled with conditions, open trades, and trades that occurred outside of regular trading hour (EST 9:30 am – 4 pm).

The information content of trade is estimated from a two-stage regression described by Foster and Viswanathan (1993). It controls for the persistence in order flow similar to models proposed by Hasbrouck (1991). We estimate the following model using intra day data:

$$Sign_{it} = a_{i0} + \sum_{j=1}^{5} a_{ij} Sign_{i,t-j} + \sum_{j=1}^{5} b_{ij} \Delta P_{i,t-j} + RSign_{it} \tag{1}$$

$$\Delta P_{it} = b_{i0} + \lambda_i * RSign_{it} + \theta_i * \Delta Sign_{it} + e_{it}, \tag{2}$$

where $Sign_{it}$ is the sign of each trade, and takes on a value of 1 for buyer-initiated trades or -1 for seller-initiated trades. $Sign_{it}$ is its jth lag. $\Delta P_{i,t-j}$ is the change in the transaction price and $\Delta P_{i,t-j}$ is its jth lag, and $RSign_{it}$ is the residual from equation (1).

In the first step, current order flow is regressed on lagged order flows and lagged price changes to get residual order flow ($RSign_{it}$). To mitigate the impact of stale information contained in the persistence of order flow and of price changes we first filter their effect out from current order flow. Following Foster and Viswanathan (1993), five lags are chosen for convenience and our results are robust to the choice of more or less lags. In the second step, price change is regressed on the residual order flow ($RSign_{it}$) from the first-stage regression, and on the difference between current trade sign and lagged trade sign. The latter captures the fixed cost component of market making described in Glosten and Harris (1988). The coefficient on the residual order flow from the second-step regression is the estimated *Lambda* (λ). This estimated *Lambda_intraday* is our proxy for the information content of trades in the intraday setting. Higher *Lambda_intraday* indicates greater information content of trades. Since Kyle's *Lambda* is defined as the sensitivity of price to order flow, we scale the estimated *Lambda* by the average daily closing price during the sample year, similar to the approach used in Brennan and Subrahmanyam (1995). This converts *Lambda* into a scale free measure and makes it comparable across stocks.

In the estimation of *Lambda*, we use trade sign rather than the size of the signed trade as a proxy for order flow based on two considerations. First, previous literature finds that signed trade size is less effective in capturing the information content of trades than the sign of the trade. Jones, Kaul and Lipson (1994) find that it is the transactions per se. rather than the size of the transactions that generates price volatility. In his study of U.S. treasury market liquidity, Fleming (2001) also finds that signed trade size is less effective at explaining price movements than the sign of the trades. Second, in our study, we link the information content of trade with the intensity of informed trading and the intensity of uniformed trading. These variables are measured as the arrival of informed trades and uninformed trades respectively defined in the structural model proposed by Easley et al. (1996) and estimated using the sign of the trades as a proxy for order flow. This allows us to maintain consistency in the measurement of all the variables of interest.

We also estimate an alternative *Lambda* (λ) measure from a similar two-step regression using daily order flows rather than intra day order flows. The following equations are estimated:

$$OIB_{it} = c_{i0} + \sum_{j=1}^{5} c_{ij} OIB_{i,t-j} + \sum_{j=1}^{5} d_{ij} R_{i,t-j} + ROIB_{it} \tag{3}$$

$$R_{it} = d_{i0} + \lambda_i * ROIB_{it} + e_{it}\ , \tag{4}$$

where OIB_{it} is daily order imbalance measured as the daily number of buyer-initiated trades minus the daily number of seller-initiated trades. $OIB_{i,t-j}$ is the *j*th lag of OIB_{it}, R_{it} is the daily return from CRSP and $R_{i,t-j}$ is its *j*th lag. $ROIB_{it}$ is the residual from the regression in equation (3). In the first step, we regress daily order imbalance on lag daily order imbalances and lag daily returns. Similar to the estimation of the intraday *Lambda*, it is important to control for the dependence of daily order imbalance on lagged daily returns and the persistence in daily order imbalance (Chordia and Subrahmanyam, 2004). Five lags are chosen for convenience and the results are not sensitive to number of lags chosen. In the second step, the daily return is regressed on the residual order imbalance (*ROIB*) from the first-step regression. The coefficient on *ROIB* is *Lambda* estimated using daily order flows (*Lambda_daily*).

2.2.3 Intensity of informed trading and uninformed trading

Estimation of informed trading and uninformed trading is based on the structural model described by Easley et al (1996). In this sequential trading model, both informed traders and uninformed traders arrive in the market following an independent Poisson process. Informed trades arrive at the rate of *Mu* (μ) only on the days with private news which occurs with probability *Alpha* (α). Informed traders buy when the news is good and sell when the news is bad. The arrival rates of uninformed buy and sell trades are *Epsilon* (ε) and are independent of the sign of the news. Thus the average number of informed trades is the product of *Alpha* and *Mu*, and that of uninformed trades is 2**Epsilon*. All these three parameters are empirically estimated using Maximum Likelihood estimation techniques.[9] We use (*Alpha***Mu*) as the proxy for the intensity of informed trading (*Informed*) and 2**Epsilon* as the proxy for the intensity of uninformed trading (*Uninformed*).[10] The relative intensity of informed trading compared with uninformed trading is the probability of information-based trading (*PIN*), which is defined as the ratio of the informed trading to the total trading,

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}. \tag{5}$$

2.2.4 Control Variables

Prior literature, specifically, Brennan and Subrahmanyam (1995), show that the information content of trades is impacted by other variables which act as control variables in a multivariate setting. We follow their analysis closely and include similar control variables described below. *Nanal* is the number of analysts following a firm, measured as the number of analysts making yearly earnings forecasts for a firm during the calendar year 2004 and estimated from the I/B/E/S Detailed History file. Trading volume is measured as the average daily number of trades (*Ntrades*) during the calendar year 2004. Return volatility (*Ret_vol*) is the standard deviation of daily CRSP close-to-close return during the calendar year 2004. *Price* is the mean of the daily closing prices from CRSP for the calendar year 2004, and is used to control for the effect of price levels

[9] See Easley et al. (1996) for the details of the model and the estimation of the model parameters.
[10] We also deflate the informed and uninformed trading by number of shareholders in the beginning of the year to reduce heteroskedasticity in the measurement of these variables. Our results remain unchanged if we use the deflated numbers.

on the information content of trades. Firm size is measured as the average of daily market capitalization (*Size*) during the calendar year 2004, calculated as the product of the daily CRSP closing price and the number of shares outstanding.

2.3 Descriptive statistics

The descriptive statistics of the variables of interest are reported in Table 3. The first column reports the cross sectional mean and standard deviation of firm characteristics for the sample used in the study. For comparison, we also report the descriptive statistics for all firms with common stocks traded on NYSE during year 2004 (referred as full sample in the following discussion). The average market capitalization (*Size*) of firms in our sample is \$ 2.92 billion, whereas, the average firm size for the full sample is \$ 6.84 billion, indicating some NYSE firms dropped due to lack of PIN are very large firms.[11] The firms in our sample are also actively traded, with an average of about 809 trades per day (*Ntrades*), however, in the full sample the average firm trades 1118 times each day. Further, average daily return volatility (*Ret_vol*) is 1.87% and for the full sample is 1.93%. The average firm in our sample has about 10 analysts making annual earnings forecast (*Nanal*), indicating that analysts actively follow the sample firms. The average firm in our sample is smaller, is traded less frequently, but has similar volatility and analyst following.

The daily number of informed trades (*Informed*) is 57 and that of uninformed trades (*Uninformed*) is 742 for the average firm. The average probability of information-based trading (*PIN*) is 10%. The mean information content of trade coefficient estimated from intraday order flows is (*Lambda_intraday*) is 5.14, which indicates that every buy (sell) transaction induces the price to increase (decrease) by about 5 basis points. The mean information content of trade coefficient estimated from daily order flow data (*Lambda_daily*) is 1.11, indicating that for every 100 net buys causes a 1.11% daily return for the average firm. The information content of trade using daily data is lower than that estimated using intraday data because the aggregation of transactions within a day in the estimation of *Lambda* reduces the information contained in each variable.

[11] The hill climbing algorithm used to calculate *PIN* does not converge for firms with very high frequency of trading.

The third and fourth columns of Table 3 report the characteristics of firms with low and high frequency of public information release. All firms are stratified into 3 equal-sized groups based on *Newsdays*, and the cross-sectional mean and standard deviation of firm characteristics within the low and high news groups are reported. Some patterns emerge from the comparison of firm characteristics across the two groups of firms. Firms with high public information release tend to have lower information content of trade. The mean information content coefficient estimated from intraday data is 7.38 for firms with low frequency of public information release, which is close to twice the information content coefficient for firms with high frequency of public information release (3.26). The same pattern holds for the information content coefficient estimated using daily data.

We also find that firms with more frequent public information release are associated with higher intensity of both informed and uninformed trading. The difference in the intensity of trading by uninformed investors between high and low public information release firms is greater than the difference in the intensity of trading by informed traders between high and low public information release firms. *Uninformed* is 383 for low public information release firms but is about one third the trading in high public information release firms (1193). The *PIN* decreases from 0.13 for low public information release firms to 0.07 for high public information release firms. A test of differences in means shows that the above differences are statistically significant.

Although the frequency of public information release is positively related with firm size, un-tabulated test results show that the differences between informed and uniformed trading are not due to differences in firm size. We stratify the firms in the sample into a 3 by 3 matrix by independently sorting into three groups each based on *Size* and *Newsdays.* We find that within each size group the firms with higher frequency of public information release are associated with, higher informed trading and higher uninformed trading, and lower *PIN* and *Lambda*, irrespective of which size group it is.

High news firms also have higher trading volume denoted by number of trades which is to be expected since both informed trading and uninformed trading is higher for such firms. Lastly, there also exists a positive relation between the frequency of public information release and the number of analysts following. This is not surprising since analysts are more likely to follow large stocks and firms covered actively in the press.

Table 4 reports the correlations between different firm characteristics. Both Pearson and Spearman (in *Italic*) correlations are reported. The frequency of public information release is positively correlated with *Informed, Uninformed*, *Ntrades*, *Nanal* and *Size*, and is negatively correlated with *PIN* and *Lambda*. We also see that *Lambda_intraday* and *Lambda_daily* are highly correlated, with a Pearson correlation coefficient of 0.81 and Spearman correlation coefficient of 0.85, which indicates that the unit of measurement either daily or intraday yields similar metrics for the information content of trades. We find that *PIN* is positively correlated with both proxies for information content of trades, suggesting that the relative intensity of informed trading versus uninformed trading plays an important role in price formation. Finally, the correlation between the *PIN* and the total trading volume is negative. This indicates that total trading is driven more by uninformed traders than informed traders.

The above univariate results suggest that firms with greater frequency of public information release are associated with lower information content of trades. Further, such firms also have greater intensity of both informed and uninformed trading.

3. Empirical results

3.1 Relation between the frequency of public information release and the information content of trade

3.1.1 Empirical model

To investigate how the information content of trade varies with the frequency of public information release, the following cross sectional regression is estimated:

$$
\begin{aligned}
Ln(Lamdba_i) = \beta_0 + \beta_1\, Ln(Newsdays_i) + \beta_2\, Ln(Nanal_i) + \beta_3\, Ln(Price_i) \\
+ \beta_4\, Ln(Ret\_Vol_i) + \beta_5\, Ln(Size_i) + \varepsilon_i. \qquad (6)
\end{aligned}
$$

Lambda is the estimated information content of trade using, either intraday data (*Lambda_intraday*), or daily data (*Lambda_daily*), defined in Section 2.2.2. *Newsdays* is the number of days with news and is a proxy for the frequency of public information release. Brennan and Subrahmanyam (1995) and Roulstone (2003) find a negative association between the information content of trade and the number of analysts following a firm. Chan and Fong (2000) find that return – order imbalance relation decreases almost monotonically with firm size. Thus, the number of analysts following (*Nanal*) a firm, and market capitalization (*Size*) are included in the regression as control variables. Other control variables, return volatility measured as the standard deviation of daily returns over the year 2004 (*Ret_Vol*) and the price level measured as the average of daily closing prices during the year 2004 (*Price*), are included to control the effect of private information flow unrelated with public information release and price level on the information content of trades, respectively.[12] Since the distribution of *Lambda* is skewed, a log-linear model rather than a linear model is estimated.[13] Specifically, all the variables are logarithm transformed and the coefficients can be interpreted as elasticities[14].

3.1.2 Results and discussion

The results of estimating Eq. (6) are reported in Table 5. The adjusted R-squares are high for both regressions, 0.82 for the regression with *Lambda_intraday* as the dependent variable and 0.69 for the regression with *Lambda_daily* as the dependent variable, indicating that the independent variables do a fair job of explaining the majority of the variation in the dependent variable. *Newsdays*, the variable we are interested in, has a coefficient of -0.10 with a t-statistic of -4.37. This indicates that every one percent increase in the number of days with news would translate into 0.1 percent decrease in the information content of trades. The negative coefficient on the frequency of public information release suggests that firms with more frequent public information release have lower information content of trades.

There are two potential explanations for the observed lower information content of trades for firms with more frequent public information release: 1) firms with frequent public information release tend to be actively traded and thus the inventory cost of trades

12 Trading volume is not included in the regression despite the previous finding that actively traded stocks are more liquid than infrequently traded ones (Stoll, 1978). The reason is that trading volume mainly captures the effect of uninformed trading as indicated by the nearly perfect correlation between the two variables. We examine the role of uninformed trading in mediating the relationship between the frequency of public information release and the information content of trade in a later section. To maintain consistency we exclude trading volume in all our analysis. However, the coefficient on Newsdays in equation (6) remains unchanged even with the addition of trading volume as a control variable.

13 All the results reported are qualitatively similar if we estimate the equations without transforming the variables into a log form.

14 Instead of using the raw values, we rank all the variables (estimate rank regressions) and our results remain unchanged.

is lower; and/or 2) public information release reduces information asymmetry, and also the relative intensity of informed trading versus uninformed trading, thus reducing the adverse selection costs. *Lambda_intraday* is estimated from trade-by-trade data and prior research suggests that inventory costs do not influence this variable while using intraday data (Stoll, 1989). Further, *Lambda_intraday* is estimated by using the sign of the orders (or net number of buyer-initiated trades when using daily data) rather than order size. Since we abstract from the size of the order, inventory costs are unlikely to impact the metric significantly and it is more likely to capture the adverse selection costs of trading.

Finally, if inventory costs drive the negative relation between the frequency of public information release and the information content of trades, then we would expect firms with frequent public information release to have relatively low levels of order imbalance and thus low inventory costs. However, un-tabulated results indicate that firms with more frequent public information release are associated with higher level of order imbalance, whether it is measured as daily order imbalance in shares (the number of buyer-initiated shares minus the number of seller-initiated shares) or the daily order imbalance in shares scaled by the number of shares outstanding. Hence, the negative relation between the frequency of public information release and the information content of trades is more likely to occur due to the reduction in adverse selection costs, which in turn is impacted by the intensity of informed trading and uninformed trading. We explore this explanation in the next section.

We see from Table 5 that the relation between the control variables and the information content of trades are consistent with findings by prior literature. Large firms, and firms actively followed by analysts are more likely to have lower information content of trades, as indicated by the significantly negative coefficients on *Nanal* and *Size*. This is consistent with results tabulated by Brennan and Subrahmanyam (1995), Roulstone (2003), and Chan and Fong (2000). The coefficient on *Ret_vol* is significantly positive, indicating that firms with more volatile return have higher information content of trades. The positive relation between return volatility and information content of trades captures the impact of other information flow unrelated to the public information release (Tauchen and Pitts, 1983). Further, the negative coefficient on *Price* indicates that high-price stocks have lower information content of trades.

The regression results with *Lambda_daily* as dependent variable are similar to those documented with *Lambda_intraday* as the dependent variable. The estimated effect of the frequency of public information release on the information content of trades is significantly negative (coefficient on *Newsdays* = -0.16). All control variables have similar signs and magnitudes as those in the regression with *Lambda_intraday* as the dependent variable, except for the coefficient on *Price* which is now positive rather than negative. Overall the regression results using daily data confirm the findings with intraday data that firms with greater frequency of public information release have lower information content of trades.

3.2 Relation between the frequency of public information release and the trading intensity of informed and uninformed traders

3.2.1 Empirical model

As discussed in section 3.1, the negative relation between the frequency of public information release and the information content of trade may partly be due to the effect of public information release on the relative intensity of informed trading versus uninformed trading. In the descriptive statistics we see that firms with higher public information release have both greater informed and uninformed trading. Further, such firms also have lower levels of relative informed trading. It is possible that the lower relative informed trading mitigates the direct effect of public information release on the information content of trades. To test this hypothesis, we first show that greater public information release results in higher levels of informed and uninformed trading. Next we introduce public information release together with informed and uninformed trading and show that the relation between the information content of trades and frequency of public information release is, at least partially, subsumed by the relation between the information content of trades and the levels of informed and uninformed trading.

We measure the relative intensity of informed trading versus uninformed trading by the probability of information-based trading (*PIN*) estimated from the structural model proposed by Easley et al (1996). *PIN* is defined as the ratio of the intensity of informed trading to total trading. Informed trading is calculated as the product of the probability of private news *Alpha* and the arrival rate of informed trades *Mu*. Total trading intensity is the sum of informed trading intensity and the intensity of uninformed trading (2**Epsilon*).

We also separately investigate the relation between the frequency of public information release and the intensity of informed trading (*Informed*) and the intensity of uninformed trading (*Uninformed*). Specifically, we estimate the following log-linear regressions:

$$Ln(Informed_i) = \gamma_0 + \gamma_1 \, Ln(Newsdays_i) + \gamma_2 \, Ln(Nanal_i) + \gamma_3 \, Ln(Size_i) + \varepsilon_i \quad (7)$$

$$Ln(Uninformed_i) = \delta_0 + \delta_1 \, Ln(Newsdays_i) + \delta_2 \, Ln(Nanal_i) + \delta_3 \, Ln(Size_i) + \varepsilon_i \quad (8)$$

$$Ln(PIN_i) = \theta_0 + \theta_1 \, Ln(Newsdays_i) + \theta_2 \, Ln(Nanal_i) + \theta_3 \, Ln(Size_i) + \varepsilon_i \quad (9)$$

Following prior literature we control for the effect of size (*Size*) and number of analysts following (*Nanal*) which have been postulated to impact the degree of information asymmetry and the preferences of different types of investors.[15] We do not include trading volume as a control variable because the *PIN* parameters are estimated directly from trading data. Theoretically, expected trading volume equals the sum of *Informed* and *Uninformed* trading, which would lead to a mechanical relation between trading volume and *PIN*.

3.2.2 Results and discussion

Table 6 shows the results of estimating Eq. (7), (8), and (9). First, we focus on the results with the intensity of informed trading, *Informed*, as the dependent variable. We note that the frequency of public information release is positively related with the intensity of informed trading. The coefficient on *Newsdays* is 0.10 and statistically significant at 1%. This result indicates that firms with more frequent public information release actually attract more informed trading. This supports the view that public information release affords sophisticated traders the opportunity to gather additional private information both before and after a public information release (Kim and Verrecchia, 1994 and 1997).[16]

The intensity of uninformed trading also increases dramatically with the frequency of public information release, as indicated by the significantly positive coefficient on *Newsdays*. This result suggests that uninformed traders prefer to trade in firms with frequent public information release. This result may seem counterintuitive

---

[15] Instead of deflating the intensity of informed and uninformed trading by the number of shareholders, we include the number of common shareholders on the RHS as an independent variable. This allows the coefficient on number of shareholders to be different from one. All our results remain the same.
16 Un-tabulated results show that the probability of private news *Alpha* increases with the frequency of public information release. The coefficient on *Newsdays* is 0.08 and the T-statistics is 4.75, when *Alpha* is the dependent variable instead of intensity of informed traders in an equation similar to Eq (7).

given the positive relation between the frequency of public information release and the intensity of informed trading. On the one hand, uninformed traders may want to avoid investing in firms where informed investors trade frequently, since, they are likely to trade against better informed investors and hence trade at an information disadvantage. On the other hand, uninformed investors more likely to be small and individual investors may not be able to gather costly private information due to constraints on resources available to them, and thus have to rely on public news to make their trading decisions. Previous studies document that small investors tend to trade after an information event, for example, an earnings announcement (Lee, 1992). Our results suggest that the latter motivation for uninformed investors to trade dominates and they do indeed trade more often in firms with higher frequency of public information releases.

The relative intensity of informed trading versus uninformed trading, measured as the probability of information-based trading *PIN*, decreases significantly with the increase in the frequency of public news. Mechanically, *PIN* is just the ratio of informed trading over total trading. Although the frequency of public information release is positively related with both informed and uninformed trading, the latter effect dominates the former and results in a negative relation between the frequency of public information release and *PIN*. *PIN* is interpreted as the probability that the market maker is likely to trade with an informed trader and thus captures the information asymmetry that the market maker faces. Our result suggests that the information asymmetry between the market maker and investors as a whole is lower for firms with more frequent public information release, and this is due to the increased trading by uninformed traders in such firms.

The number of analysts following, which also captures the information production activity of investors, is positively related with both the intensity of informed trading and that of uninformed trading, but is negatively related with *PIN*. Brennan and Subrahmanyam (1995) find a negative relation between the number of analysts following and the information content of trades. They ascribe a private information production role to analysts. Roulstone (2003) confirms this finding but interprets it differently. He suggests that analysts' reports are publicly available and so analysts should be viewed as providing public news to the market rather than private information. Our findings seem to suggest that analysts are a proxy for private information because of the positive relation with informed trading and also act as public information providers due to the positive relation with uninformed trading.

3.3 The role of the intensity of informed trading and uninformed trading on the relation between public information release and the information content of trade

3.3.1 Empirical model

In the sections above we first show that firms with greater frequency of public information release have lower information content of trades on average. We conjecture that this negative relationship could be due to the impact of public information release on informed and uninformed trading. We next show that firms with greater frequency of public information release have greater informed trading as well as greater uninformed trading. We put the above two results together and test whether greater intensity of informed and uninformed trading in stocks with higher public information release accounts for all or part of the relation between public information release and information content of trade. We estimate the following log-linear model,

$$Ln(Lamdba_i) = \kappa_0 + \kappa_1\, Ln(Newsdays_i) + \kappa_2\, Ln(Informed_i) + \kappa_3\, Ln(Uninformed_i) + \kappa_4\, Ln(Nanal_i) + \kappa_5\, Ln(Price_i) + \kappa_6\, Ln(Ret\_Vol_i) + \kappa_7\, Ln(Size_i) + \varepsilon_i \quad (10)$$

Similar to our previous analysis *Lambda* is either *Lambda_intraday* or *Lambda_intraday*, which are calculated using intra day data and daily data, respectively. *Informed* is the proxy for the intensity of informed trading, measured as the product of the probability of private news and the arrival rate of informed trades, and *Uninformed* is the proxy for the intensity of uninformed trading, measured as the arrival rate of uninformed trades.

We expect that after controlling for the role of *Informed*, and *Uninformed* the effect of *Newsdays* on *Lambda* is likely to be less significant than previously documented. Our inference from the mediation effect of informed and uninformed trading on the relation between *Lambda* and frequency of public information release is similar to analysis common in the social sciences (Baron and Kenny, 1986). Since *PIN* is the ratio of the intensity of informed to total trading, we estimate equation (11) in which we substitute the comprehensive measure *PIN* for the two components, *Informed* and *Uninformed*.

$$Ln(Lamdba_{\mathrm{i}}) = \pi_0 + \pi_1\, Ln(Newsdays_{\mathrm{i}}) + \pi_2\, Ln(PIN_{\mathrm{i}}) + \pi_3\, Ln(Nanal_{\mathrm{i}})$$
$$+ \pi_4\, Ln(Price_{\mathrm{i}}) + \pi_5\, Ln(Ret\_Vol_{\mathrm{i}}) + \pi_6\, Ln(Size_{\mathrm{i}}) + \varepsilon_{\mathrm{i}} \qquad (11)$$

3.3.2 Result and discussion

The results from estimating Eq. (10) and (11) are reported in Table 7. Panel A presents results relating to the mediating role of informed trading and uninformed trading. We present results which are estimated with *Lambda_intraday* as the dependent variable, and note that results using *Lambda_daily* as the dependent variable remain the same. We find that the coefficient on *Informed* is 0.18 and is significant at 1% level. This indicates that the intensity of informed trading significantly increase the information content of trades. Ceteris paribus, increased informed trading makes the market maker more likely to lose money by trading against better informed agents, which causes the order flow to contain greater amount of price relevant information. We also find that the trading by uninformed traders decreases the information content of trade, as indicated by the significantly negative coefficient on *Uninformed*. Interestingly, *Newsdays* is no longer significant after controlling for *Informed* and *Uninformed*, although it is still negatively related to *Lambda*. This result suggests that almost all of the effect of public information release on the information content of trades is due to the mediation of informed trading and uninformed trading.

The result using *PIN*, which is reported in Panel B of Table 7, confirms the result using *Informed* and *Uninformed* discussed above. First, *PIN* is significantly positively related to *Lambda*, which indicates the order flow contains greater price relevant information when the probability of a trade emanating from an informed trader is higher. Second, *Newsdays* is again not significant after controlling for *PIN*, suggesting that its effect on the information content of trades is completely subsumed by the effect of *PIN* on the information content of trades.

3.4 Possible endogenous relation between Lamdba and orderflow.

Brennan and Subrahmanyam (1995) suggest that *Lambda* and order flow could be endogenously determined (page 369). To control for a possible endogenous relation between these two variables we estimate the following equations using two stage least squares:

$$Ln(Lamdba_i) = \omega_0 + \omega_1\, Ln(Newsdays_i) + \omega_2\, Ln(PIN_i) + \omega_3\, Ln(Price_i) + \omega_4\, Ln(Ret\_Vol_i) + \omega_5\, Ln(Size_i) + \omega_6\, Ln(LagLamdba_i) + \varepsilon_i \qquad (12)$$

$$Ln(PIN_i) = \varphi_0 + \varphi_1\, Ln(Newsdays_i) + \varphi_2\, Ln(Lambda_i) + \varphi_3\, Ln(Size_i) + \varphi_4\, Ln(LagPIN_i) + \varphi_5\, Ln(Institutions_i) + \varepsilon_i \qquad (13)$$

In equation (12) the price of the stock (*Price*) and the Lag of Lambda (*LagLambda*) identify the *Lambda* equation. Two variables help identify the *PIN* equation. These are lagged values of *PIN* (*LagPIN*) and the percentage of institutional ownership (*Institutions*).

Untabulated results show that the frequency of public news release is positively (but not significantly) related to *Lambda*. Further, *PIN* is significantly positively related to *Lambda*, which suggests that results documented in the main text of the paper are robust to controlling for a possible endogenous relationship between *Lambda* and *PIN.* The results from estimating the second equation shows that the frequency of news releases is significantly negatively related with *PIN*, confirming results documented using single equation models.

4. Conclusion

Order flow plays an important role in price discovery and this topic has been of interest to researchers in finance. Theory suggests that the information content of trades is dependent on the information advantage informed traders have over the market maker who is assumed to be uninformed. Further, other papers suggest that public information can increase the information content of trades by inducing informed traders to gather private information. However, public information release is also likely to attract uninformed traders because they use these releases to form their trading strategy. This is likely to decrease the information content of trades for firms with greater public information release. Our results show that greater frequency of public information release reduces the information content of trades. Next, we show that greater public information release increases both informed and uninformed trading though the latter effect dominates. Finally we show that the lower information content of trades for firms with higher public disclosure is driven by the greater trading by uninformed traders in such stocks. Our results support the emphasis that the SEC has placed on greater public information release. This helps reduce the adverse selection costs in the market, increase liquidity in the stock, and level the playing field for all investors.

References:

Bagnoli, M., Watts, S., 1998. Information acquisition, information release and trading dynamics. Journal of Financial Markets 1, 221 – 252.

Barber, B., Odean, T., 2005. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. Unpublished working paper. University of California, Berkeley.

Baron, R., Kenny, D., 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 51, 1173-1182.

Brennan, M., Subrahmanyam, A., 1995. Investment analysis and price formation in securities markets. Journal of Financial Economics 38, 361-381.

Chan, K., Fong, W., 2000. Trade size, order imbalance, and the volatility volume relation. Journal of Financial Economics 57, 247-273.

Chordia, T., Subrahmanyam, A., 2004. Order imbalance and individual stock returns: theory and evidence. Journal of Financial Economics 72, 485-518.

Cong, Y., Hoitash, R., Krishnan, M., 2005. Event study with imperfect competition and private information: earnings announcements revisited. Unpublished working paper. Baruch College.

Diamond, D., Verrecchia, R., 1991. Disclosure, liquidity, and the cost of capital. Journal of Finance 46, 1325–1359.

Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns?. Journal of Finance 57, 2185–2221.

Easley, D., Hvidkjaer, S., O'Hara, M., 2005. Factoring information into returns. Unpublished working paper. University of Maryland.

Easley, D., Kiefer, N., O'Hara, M., Paperman, J., 1996. Liquidity, information, and infrequently traded stocks. Journal of Finance 51, 1405–1436.

Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69 – 90.

Easley, D., O'Hara, M., 2004. Information and the cost of capital. Journal of Finance 59, 1553 – 1583.

Fleming, M., 2001. Measuring treasury market liquidity. Staff Report, Federal Reserve Bank of New York.

Foster, D., Viswanathan, S., 1993. Variations in trading volume, return volatility, and trading costs: evidence on recent price formation models. Journal of Finance 48, 187-211.

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

Glosten, L., Milgrom, P., 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71-100.

Graham, J., Koski, J., Loewenstein, U., 2006. Information flow and liquidity around anticipated and unanticipated dividend announcements. Journal of Business, forthcoming.

Grundy, B., McNichols, M., 1989. Trade and the revelation of information through prices and direct disclosure. Review of Financial Studies 2, 495–526.

Hasbrouck, J., 1991. Measuring the information content of stock trades. Journal of Finance 46, 179-207.
Jones, C., Kaul, G., Lipson, M., 1994. Transactions, volume and volatility. Review of Financial Studies 7, 631–651.
Kim, O., Verrecchia, R., 1994. Market liquidity and volume around earnings announcements. Journal of Accounting and Economics 17, 41–67.
Kim, O., Verrecchia, R., 1997. Pre-announcement and event-period private information. Journal of Accounting and Economics 24, 395–419.
Koski, J., Michaely, R., 2000. Prices, liquidity, and the information content of trades. Review of Financial Studies 13, 659–696.
Krinsky, I., Lee, J., 1996. Earnings announcements and the components of the bid-ask spread. Journal of Finance 51, 1523–1535.
Kyle, A., 1985. Continuous auctions and insider trading. Econometrica 53, 1315-1335.
Lee, C., 1992. Earnings news and small traders: an intraday analysis. Journal of Accounting and Economics 15, 265-302.
Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733-746.
Livne, G., 2000. Information asymmetry, investment horizons and the dual role of public announcements. Review of Accounting Studies 5, 127-153.
Pastor, L., Stambaugh, R., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642 – 685.
Ranaldo, A., 2004. Intraday market dynamics around public information arrivals. Unpublished working paper. University of Fribourg.
Roulstone, D., 2003. Analyst following and market liquidity. Contemporary Accounting Research 20, 551 – 578.
Stoll, H., 1978. The pricing of dealer services: an empirical study of NASDAQ stocks. Journal of Finance 33, 1152-1173.
Stoll, H., 1989. Inferring the components of the bid-ask spread. Journal of Finance 44, 115-134.
Tauchen, G., Pitts, M., 1983. The price variability-volume relationship on speculative markets. Econometrica 51, 485-505.
Verrecchia, R., 2001. Essays on disclosure. Journal of Accounting and Economics 32, 97–180.

Table 1
Definition of variables used in the analyses

Variable	Definition
Newsdays	The number of days with at least one public information release on www.MarketWatch.com
Informed	The number of daily informed trades in hundreds, measured as the product of the probability of private news and the arrival rate of informed trades which are estimated from the model proposed by Easley et al. (1996)
Uninformed	The number of daily uninformed trades in hundreds, measured as the arrival rate of uninformed trades which is estimated from the model proposed by Easley et al. (1996)
PIN	The relative intensity of informed trading versus uninformed trading, measured as the probability of information-based trading, i.e., Informed/(Informed + Uninformed), which is estimated from the model proposed by Easley et al. (1996)
*Lambda_intraday**	The information content of trade estimated from a two-step regression as specified in Eq. (1) and (2) using intraday data
*Lambda_daily***	The information content of trade estimated from a two-step regression as specified in Eq. (3) and (4) using daily data
Nanal	The number of analysts making yearly earnings forecasts for a firm during year 2004 and included in I/B/E/S detail history file
Ntrades	The average daily number of trades during year 2004
Ret_Vol	The standard deviation of daily return during year 2004
Price	The average of daily closing price during year 2004
Size	The average market capitalization during year 2004

Table 1 Continued…

* *Lambda_intraday* is estimated from the following two-step regression,

$$Sign_{it} = a_{i0} + \sum_{j=1}^{5} a_{ij} Sign_{i,t-j} + \sum_{j=1}^{5} b_{ij} \Delta P_{i,t-j} + RSign_{it} \tag{1}$$

$$\Delta P_{it} = b_{i0} + \lambda_i * RSign_{it} + \theta_i * \Delta Sign_{it} + e_{it}, \tag{2}$$

where $Sign_{it}$ is the sign of the trade which takes value 1 for buyer-initiated trades or -1 for seller-initiated trades and $Sign_{i,t-j}$ is its j[th] lag, ΔP_{it} is change in transaction price and $\Delta P_{i,t-j}$ is its j[th] lag, and $RSign_{it}$ is the residual from the regression in equation (1). *Lambda_intraday* is the estimated coefficient on $RSign_{it}$ scaled by the average daily closing price.

** *Lambda_daily* is estimated from the following two-step regression,

$$OIB_{it} = c_{i0} + \sum_{j=1}^{5} c_{ij} OIB_{i,t-j} + \sum_{j=1}^{5} d_{ij} R_{i,t-j} + ROIB_{it} \tag{3}$$

$$R_{it} = d_{i0} + \lambda_i * ROIB_{it} + e_{it}, \tag{4}$$

where OIB_{it} is daily order imbalance measured as the daily number of buyer-initiated trades minus the daily number of seller-initiated trades and $OIB_{i,t-j}$ is its jth lag, R_{it} is daily return from CRSP and $R_{i,t-j}$ is its jth lag, and $ROIB_{it}$ is the residual from the regression in equation (3). *Lambda_daily* is the estimated coefficient on $ROIB_{it}$.

Table 2
Distribution of news for common stocks traded on the NYSE

This table describes the distribution of news releases appearing on www.MarketWatch.com during the year 2004 for NYSE common stocks. Out of 1535 NYSE common stocks in CRSP database, 1362 stocks are covered in www.MarketWatch.com during year 2004. *Newsdays* is a cross sectional distribution of the average for each firm of the number of days with at least one news release. *Yearnews* is the cross sectional distribution of the average for each firm of the total number of news release during the whole year.

	News variables	
Statistics	*Newsdays*	*Yearnews*
Mean	70.22	177.22
Std.	47.32	276.60
Min	7	11
Q1 (25%)	39	67
Median (50%)	55	98
Q3 (75%)	85	170
Max	252	3456

Table 3
Descriptive statistics

This table presents the descriptive statistics for a sample of 1028 firms during 2004. The first column presents the descriptive statistics for the firms with public news. For comparison, the descriptive statistics for all NYSE common stocks is reported in the second column. The next three columns compare the descriptive statistics of firms with different frequency of public information release. Firms in the sample are first stratified into three equally-sized groups, specifically, low news firms, medium news firms and high news firms, according to the frequency of public information release. The descriptive statistics of low news firms and high news firms are reported. The descriptive statistics of medium news firms are not reported for easier readability. T-statistics of the differences in mean value between high news firms and low news firms are reported in the last column. See Table 1 for the definitions of variables.

	With news firms (N = 1028)		NYSE firms (N = 1535)		Low news firms (N = 348)		High news firms (N = 352)		T-stat
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	(High – Low)
Newsdays	59.54	32.85	NA	NA	33.31	6.74	93.67	34.51	32.03
Informed (100s)	0.57	0.25	NA	NA	0.43	0.21	0.73	0.24	18.24
Uninformed (100s)	7.42	5.60	NA	NA	3.83	3.05	11.93	6.01	22.46
PIN	0.10	0.05	NA	NA	0.13	0.05	0.07	0.03	-19.97
Lambda_intraday (% per 100 trades)	5.14	4.17	5.25	4.99	7.38	5.03	3.26	2.49	-13.69
Lambda_daily (% per 100 trades)	1.11	1.44	1.28	2.08	1.77	1.94	0.57	0.42	-11.33
Nanal	10.19	7.19	11.22	9.44	5.99	4.59	15.08	7.52	19.27
Ntrades (100s)	8.09	5.93	11.18	12.91	4.28	3.27	12.85	6.34	22.44
Ret_Vol (%)	1.87	0.72	1.93	0.88	1.96	0.73	1.77	0.74	-3.38
Price	31.72	34.55	91.60	22.69	28.06	24.53	36.14	51.39	2.65
Size ($Billion)	2.92	4.69	6.84	21.63	1.17	1.54	5.59	6.78	11.87

Table 4
Correlation coefficients between firm characteristics

This table reports the correlations between firm characteristics. Pearson correlations are reported in the upper-right section and Spearman correlations are reported in the lower-left section in *Italic*. See Table 1 for the definitions of variables.

	Newsdays	*Informed*	*Uninformed*	*PIN*	*Lambda_ intraday*	*Lambda_ Daily*	*Nanal*	*Ntrades*	*Ret_Vol*	*Size*
Newsdays		0.54	0.65	-0.47	-0.36	-0.30	0.53	0.65	-0.14	0.56
Informed	*0.54*		0.85	-0.62	-0.66	-0.55	0.64	0.86	-0.32	0.65
Uninformed	*0.66*	*0.87*		-0.73	-0.61	-0.46	0.74	0.99	-0.26	0.76
PIN	*-0.63*	*-0.60*	*-0.90*		0.81	0.70	-0.57	-0.73	0.13	-0.40
Lambda_ intraday	*-0.56*	*-0.77*	*-0.87*	*0.79*		0.81	-0.53	-0.62	0.49	-0.39
Lambda_ daily	*-0.54*	*-0.70*	*-0.77*	*0.70*	*0.85*		-0.40	-0.46	0.33	-0.28
Nanal	*0.57*	*0.67*	*0.75*	*-0.67*	*-0.71*	*-0.66*		0.74	-0.28	0.55
Ntrades	*0.66*	*0.88*	*0.99*	*-0.90*	*-0.87*	*-0.77*	*0.75*		-0.26	0.76
Ret_Vol	*-0.13*	*-0.32*	*-0.26*	*0.14*	*0.59*	*0.58*	*-0.30*	*-0.26*		-0.41
Size	*0.53*	*0.74*	*0.81*	*-0.72*	*-0.92*	*-0.82*	*0.70*	*0.82*	*-0.62*	

Table 5
Relation between the frequency of public information release and the information content of trades

This table presents the result of regressing the information content of trades on the frequency of public information release, measured as number of days with news appearing on www.MarketWatch.com during year 2004 (*Newsdays*), for a sample of 1028 firms during the sample year 2004. The regression equation is specified in Eq. (6),

$$Ln(Lamdba_i) = \beta_0 + \beta_1\, Ln(Newsdays_i) + \beta_2\, Ln(Nanal_i) + \beta_3\, Ln(Price_i) + \beta_4\, Ln(Ret\_Vol_i) + \beta_5\, Ln(Size_i) + \varepsilon_i. \tag{6}$$

where the dependent variables *Lambda_intraday* and *Lambda_daily* which are modified Kyle's *Lambda* are estimated using intraday data and daily data respectively. See Table 1 for the definitions of variables. All variables are logarithm transformed and OLS is used to estimate the coefficients.

	Dependent variable	
	Ln(Lambda_intraday)	*Ln(Lambda_daily)*
Independent Variable	Coefficient (T-value)	Coefficient (T-value)
Intercept	2.70*** (23.25)	0.11 (0.59)
Ln(Newsdays)	-0.10*** (-4.37)	-0.16*** (-4.12)
Ln(Nanal)	-0.12*** (-8.30)	-0.19*** (-8.05)
Ln(Price)	-0.17*** (-9.98)	0.17*** (5.76)
Ln(Ret_Vol)	0.07** (1.98)	0.43*** (7.76)
Ln(Size)	-0.34*** (-24.71)	-0.40*** (-17.16)
Adj. R^2	0.82	0.69

, **, ** indicate significance at the 10%, 5%, and 1% level, for a two tailed test, respectively.

Table 6
Relation between the frequency of public information release and the intensity of informed and uninformed trading

This table presents the result of regressing the intensity of informed and uninformed trading on the frequency of public information release using a sample of 1028 NYSE common stocks in sample year 2004. The regression equations are Eq. (7) – (9),

$$Ln(Informed_i) = \gamma_0 + \gamma_1\, Ln(Newsdays_i) + \gamma_2\, Ln(Nanal_i) + \gamma_3\, Ln(Size_i) + \varepsilon_i \qquad (7)$$

$$Ln(Uninformed_i) = \delta_0 + \delta_1\, Ln(Newsdays_i) + \delta_2\, Ln(Nanal_i) + \delta_3\, Ln(Size_i) + \varepsilon_i \qquad (8)$$

$$Ln(PIN_i) = \theta_0 + \theta_1\, Ln(Newsdays_i) + \theta_2\, Ln(Nanal_i) + \theta_3\, Ln(Size_i) + \varepsilon_i \qquad (9)$$

The frequency of public information release (*Newsdays*) is measured as number of days with news appearing on www.MarketWatch.com during year 2004 for each firm. The intensity of informed trading (*Informed*) is measured as the product of the probability of private news occurrence and the arrival rate of informed trades, and intensity of uninformed trading (*Uninformed*) is measured as the arrival rate of uninformed trades. *PIN* is the probability of information-based trading, which measures the relative intensity of informed trading versus uninformed trading and is calculated as the ratio of *Informed* to the sum of *Informed* and *Uninformed* trades. All three parameters, the probability of private news, the arrival rate of informed trades, and the arrival rate of uninformed trades, are estimated according to the model proposed by Easley et al (1996). See Table 1 for the definitions of variables. All variables are logarithm transformed and OLS is used to estimate the coefficients.

	Dependent variable		
	Ln(Informed)	*Ln(Uninformed)*	*Ln(PIN)*
Independent Variable	Coefficient (T-value)	Coefficient (T-value)	Coefficient (T-value)
Intercept	-1.47*** (-13.34)	-0.43*** (-2.68)	-1.26*** (-14.43)
Ln(Newsdays)	0.10*** (3.32)	0.33*** (7.87)	-0.22*** (-9.50)
Ln(Nanal)	0.16*** (8.52)	0.30*** (11.18)	-0.13*** (-8.64)
Ln(Size)	0.24*** (18.35)	0.43*** (22.24)	-0.17*** (-16.18)
Adj. R^2	0.58	0.71	0.61

, **, ** indicate significance at the 10%, 5%, and 1% level, for a two tailed test, respectively.

Table 7
Relation between the frequency of public information release, the intensity of informed trading, uninformed trading, and the information content of trades

This table presents the result of regressing the information content of trades on the frequency of public information release (*Newsdays*), the intensity of informed trading (*Informed*) and the intensity of uninformed trading (*Uninformed*), for a sample of 1028 firms during 2004. Panel A reports the result of the regression with *Informed* and *Uninformed* as regressors as specified in Eq. (10) and panel B reports the result of the regression using *PIN* as regressor as specified in Eq. (11),

$$Ln(Lamdba_i) = \kappa_0 + \kappa_1\, Ln(Newsdays_i) + \kappa_2\, Ln(Informed_i) + \kappa_3\, Ln(Uninformed_i) + \kappa_4\, Ln(Nanal_i) + \kappa_5\, Ln(Price_i) + \kappa_6\, Ln(Ret\_Vol_i) + \kappa_7\, Ln(Size_i) + \varepsilon_i \quad (10)$$

$$Ln(Lamdba_i) = \pi_0 + \pi_1\, Ln(Newsdays_i) + \pi_2\, Ln(PIN_i) + \pi_3\, Ln(Nanal_i) + \pi_4\, Ln(Price_i) + \pi_5\, Ln(Ret\_Vol_i) + \pi_6\, Ln(Size_i) + \varepsilon_i \quad (11)$$

The dependent variables *Lambda_intraday* and *Lambda_daily* which are modified Kyle's *Lambda* are estimated using intraday data and daily data respectively. The frequency of public information release (*Newsdays*) is measured as number of days with news appearing on www.MarketWatch.com during year 2004 for each firm. The intensity of informed trading (*Informed*) is measured as the product of the probability of private news and the arrival rate of informed trades, and intensity of uninformed trading (*Uninformed*) is measured as the arrival rate of uninformed trades. *PIN* is the probability of information-based trading, which measures the relative intensity of informed trading to toal trading and is calculated as the ratio of *Informed* to the sum of *Informed* and *Uninformed* trading. All three parameters, the probability of private news, the arrival rate of informed trades, and the arrival rate of uninformed trades, are estimated according to the model proposed by Easley et al (1996). See Table 1 for the definitions of variables. All variables are logarithm transformed and OLS is used to estimate the coefficients.

Table 7 Continued….

Panel A: Regression with *Informed* and *Uninformed* as independent variables

Independent Variable	Dependent variable: *Ln(Lambda_intraday)* Coefficient (T-value)	Dependent variable: *Ln(Lambda_daily)* Coefficient (T-value)
Intercept	2.80*** (33.91)	0.29* (1.77)
Ln(Newsdays)	-0.02 (-0.94)	-0.04 (-1.21)
Ln(Informed)	0.18*** (5.59)	0.30*** (5.02)
Ln(Uninformed)	-0.55*** (-24.52)	-0.77*** (-17.56)
Ln(Nanal)	-0.02 (-1.51)	-0.05*** (-2.62)
Ln(Price)	-0.17*** (-13.84)	0.18*** (7.29)
Ln(Ret_Vol)	0.52*** (19.20)	1.07*** (20.02)
Ln(Size)	-0.03*** (-2.46)	0.03 (1.06)
Adj. R^2	0.92	0.80

Panel B Regression using *PIN*

Independent Variable	Dependent variable: *Ln(Lambda_intraday)* Coefficient (T-value)	Dependent variable: *Ln(Lambda_daily)* Coefficient (T-value)
Intercept	3.33*** (33.34)	1.00*** (5.59)
Ln(Newsdays)	0.00 (-0.19)	-0.03 (-0.77)
Ln(PIN)	0.65*** (21.91)	0.90*** (16.67)
Ln(Nanal)	-0.06*** (-5.16)	-0.12*** (-5.33)
Ln(Price)	-0.12*** (-8.05)	0.24*** (9.25)
Ln(Ret_Vol)	0.44*** (13.76)	0.95*** (16.29)
Ln(Size)	-0.16*** (-10.91)	-0.14*** (-5.46)
Adj. R^2	0.88	0.76

, **, ** indicate significance at the 10%, 5%, and 1% level, on a two tailed test, respectively.