AIMing at PIN: Order Flow, Information, and Liquidity*

Gautam Kaul, Qin Lei and Noah Stoffman

July 16, 2008

ABSTRACT

In this study, we model and measure the existence of informed trading. Specifically, we investigate the properties of the widely used measure of informed trading, PIN, developed by Easley and O'Hara, and establish three important features of informed trading. First, the existence of informed trading, and therefore PIN, should be estimated over different trading intervals for stocks of different characteristics. Second, we establish a direct relationship between PIN and the absolute (percentage) order imbalance (AIM). The latter is not only easier to measure, but can also be readily calculated over short horizons. Most importantly, we show that conditions for the theoretical equivalence between estimated PIN and AIM of a stock serve as a guide for the optimal estimation interval that should be used for that particular stock. Finally, and significantly, an investigation around exogenous national security events reveals strong evidence against interpreting PIN and order imbalance as a liquidity measure.

Keywords: Stock Liquidity; Information Asymmetry; Information Content; Order Imbalance; PIN

---

*Gautam Kaul is at the Finance Department of Stephen M. Ross School of Business at the University of Michigan; Qin Lei is at the Finance Department of Edwin L. Cox School of Business at Southern Methodist University; and Noah Stoffman is at the Finance Department of Kelley School of Business, Indiana University. We thank Hank Bessembinder, Nick Bollen, Alex Butler, Craig Holden, Aditya Kaul, Ron Masulis, Hans Stoll, Charles Trzcinka, and Kumar Venkataraman for very helpful discussions. Address correspondence to Qin Lei via email at *qlei@cox.smu.edu*.

Following the seminal work of Akerlof (1970), Spence (1974), and Rothschild and Stiglitz (1976), information asymmetry has occupied a central role in financial economics. In a series of influential papers, Easley and O'Hara, along with their co-authors, popularize the concept of "probability of informed trading" (also known as PIN). The theoretical basis and empirical evidence in several papers justify interpreting PIN as a good proxy for information asymmetry, and there is also evidence to support the existence of an information risk factor rooted in the estimated PINs.[1] These conclusions however have recently received a healthy dose of scrutiny, with most of the criticism centered around whether PIN measures information asymmetry or illiquidity.[2]

In this paper, we examine the micro-foundation of the PIN measure with three specific goals in mind. First, we demonstrate that the news arrival process of a stock is inherently peculiar to its characteristics and should, in turn, determine both the modeling and estimation of PIN. Specifically, we argue and demonstrate that the PIN measure cannot be estimated using the same trading interval for stocks with different informational characteristics. The widespread use of a fixed (typically daily) interval to estimate PINs of all stocks is likely to lead to systematic estimation errors in PINs of all stocks. Such errors can, in turn, lead to erroneous conclusions about the existence, extent, and behavior of asymmetric information across different stocks. Second, under reasonable conditions that are consistent with the theoretical modeling of PIN, we show that order imbalances are a legitimate proxy for information asymmetry. We derive a positive and linear relationship between the PIN measure and absolute (percentage) order imbalances (hereafter, AIM). The main advantage of this result is that AIM is significantly easier to compute and, more importantly, unlike PIN, can be calculated for very short trading intervals when information asymmetry is likely to be prevalent. Importantly, we also use the equivalence between PIN and AIM to provide guidelines for the optimal trading interval to estimate the extent of information asymmetry for stocks of different informational characteristics.

Finally, we conduct an experiment to determine whether PIN reflects liquidity, instead of

[1]See, for example, Easley et al. (1996), Easley et al. (1998), Easley, O'Hara and Saar (2001), Easley et al. (2002, 2005), Brown et al. (2004), Easley and O'Hara (2004), Vega (2006) and Duarte et al. (2008).

[2]See, for example, Spiegel and Wang (2005), Mohanram and Rajgopal (2006), Aktas et al. (2007), Boehmer et al. (2007) and Duarte and Young (2007).

information asymmetry, as argued in several recent papers. Specifically, we evaluate whether estimates of AIM, and therefore PIN, change in an environment in which liquidity is likely to be exogenously and unexpectedly affected, but information asymmetry is unlikely to change. We examine the change in these measures around exogenous and unpredictable events such as the attack on the World Trade Center in 2001. During such an event we would expect a drop in liquidity, but no change in information asymmetry. We find no evidence in support of the hypothesis that PIN, and therefore order imbalances, reflect reduced liquidity. Specifically, while we find a significant and widespread drop in standard measures of liquidity, there is no change in the estimates of AIM around these events.

The rest of the paper is organized as follows. In Section I, we present a simple model of the trading process to show the explicit functional relation between the fraction of informed orders and the absolute order imbalance. In Section II, we present empirical evidence that the behavior of PIN is dependent both on the length of sampling interval and, more importantly, on the cross-sectional characteristics of different stocks. We also empirically establish the relation between absolute order imbalances and PIN and, importantly, use this relationship to determine the different optimal sampling intervals for measuring the extent of asymmetric information for different stocks. In Section III we provide evidence that the PIN and AIM measures do not reflect illiquidity. Section IV contains a summary and some concluding remarks.

I. Model

Consider a trading process in which the market maker facilitates orders for the stock of a specific firm. Investors are classified into two groups. Informed investors either have access to private information regarding the true value of the underlying firm or possess superior skills at processing public information. Uninformed investors, on the other hand, do not possess any such informational or information-processing skills. The market maker observes aggregate order flow and tries to infer the information content of the orders submitted by both types of traders.

Assume that informed and uninformed orders follow two independent Poisson processes with expected arrival rates ι and ν over any particular time interval Δt, respectively. The trading process of the informed and uninformed differs which allows Easley and O'Hara to derive the probability of informed trading measure, also known as PIN. The uninformed orders could be buy/sell orders with equal probability 0.50, while the informed orders always take only one side of the trade.

The PIN measure therefore indirectly achieves the separation between informed and uninformed trading because informed traders are assumed to always trade on the right side (buying on good news and selling on bad news), whereas uninformed traders submit buy and sell orders with equal probabilities. These set of structural assumptions collectively lead to the empirical identification of order arrival rates and other parameters necessary for the calculation of the PIN variable.

The presence of an order imbalance (defined as buy orders in excess of sell orders) provides evidence of informed trading activities within this trading framework. When there is good news, more informed buy orders imply a larger positive order imbalance. When there is bad news, more informed sell orders imply a larger negative order imbalance. Either way, a higher fraction of informed trades should be associated with a higher (percentage) absolute order imbalance. We now focus on developing the exact functional relationship between probability of informed trading and absolute order imbalance.

Assume that conditional on the submission of an informed order, the probability of an informed buy order is ϕ [and therefore the probability of an informed sell order is $(1 - \phi)$]. Given rational expectations, the buy orders over the time interval Δt are given by $B = \phi\iota + (1/2)\nu$, and the sell orders are $S = (1 - \phi)\iota + (1/2)\nu$. The fraction of informed orders consequently is $PIN = \iota/(\iota + \nu)$.

Although the derivation of PIN appears quite simple and intuitive, its estimation is complex. This is because we need to solve for the parameters $\{\phi, \iota, \nu\}$ using the order flows $\{B, S\}$. Clearly, this requires a researcher to have one additional degree of freedom to uniquely determine the parameter values and, hence, the instantaneous fraction of informed trades given the order flows observed over the particular time interval.

Easley et al. (1996) construct a statistical model of order arrival processes and rely on the statistical properties of the aggregate order flows to infer the information content of trades. The essential advantage of this novel approach is to avoid identifying informed traders at each transaction. Specifically, it is assumed that the uninformed traders submit both buy and sell orders at the same rate, while the informed traders submit orders only on the side consistent with the information they receive (i.e., buy upon good news or sell upon bad news). The one-sided nature of the informed trades is the key to the empirical distinction between the informed and the uninformed orders, and the identification is achieved through the likelihood of observing a given stream of orders corresponding to different combinations of order arrival rates for the informed and uninformed traders. The combination of arrival rates that generates the highest sample likelihood is used to compute the most-likely fraction of informed orders. This measure is known as the probability of informed trading, or PIN.

The PIN measure is difficult to estimate because the order arrival rates have to be estimated using a maximum likelihood procedure that is time consuming and sometimes numerically unstable.[3] Moreover, since estimated PINs are typically available at the quarterly intervals, but not at a higher frequency, PIN is not well-suited to studying information asymmetry when it is most likely to be prevalent during short-lived corporate events, such as earnings or merger announcements.

Fortunately, however, one can deal directly with the fraction of informed trades. As shown in the appendix, there exists a one-to-one correspondence between the fraction of informed orders, PIN, and the signed (percentage) order imbalance $SIM \equiv (B - S)/(B + S)$.

Proposition 1 *Over any specific trading interval, Δt, the signed (percentage) order imbalance is proportional to the fraction of informed orders, SIM, that is, $SIM = (2\phi - 1) \cdot PIN$.*

Proof. See Appendix. ■

It is intuitive that the type of news released to, or inferred by, informed traders determines the sign of proportionality $(2\phi - 1)$ which, in turn, conforms to the direction of order im-

[3] The estimation is sensitive to the initial values used to start the maximum-likelihood procedure. It sometimes breaks down and stops at a local maximum instead of a desired global maximum. See Yan and Zhang (2006) for a more detailed discussion.

balances. Put differently, the informed traders create positive (or negative) order imbalances during a trading interval Δt that is perceived to be associated with good (or bad) news. Without a good handle of the news probability ϕ, however, it is hard to make use of this intuitive relationship. Specifically, the cross-sectional comparison of signed order imbalances for different firms is not necessarily reflective of the fraction of informed orders, and the time-series comparison of signed order imbalances for the same firm need not be consistent with changes in information content.

At the cost of some more structure on the trading process, however, a more useful relationship between PIN and order imbalances can be determined. For simplicity, and again consistent with the microstructure literature on PIN assume that informed traders do not trade strategically;[4] they never randomize their orders to camouflage their intended trading activities. As a consequence, consecutive orders from informed investors will be of the same type (either buy or sell) until the type of news reverses, and ϕ will merely oscillate between 0 and 1. Specifically, ϕ takes the value of 1 during a trading interval that is characterized by the arrival of good news, and takes the value of 0 during an interval that experiences bad news. During a period of no news arrival, the informed traders will not submit orders, resulting in a drop in the informed arrival rate. Over a calendar time period covering multiple trading intervals $s = k \cdot \Delta t$, where k is a positive integer, the mean value of ϕ will measure the fraction of time when the underlying firm reveals good news.

Under these assumptions, there exists an equivalence between the fraction of informed orders PIN and the absolute (percentage) order imbalance $AIM \equiv |(B-S)/(B+S)|$.

Proposition 2 *During a trading interval Δt in which there is no news reversal, and assuming that informed investors always submit orders conforming to the type of news they receive, the absolute (percentage) order imbalance is equivalent to the fraction of informed orders, i.e., $AIM = PIN$.*

Proof. Since ϕ oscillates between 0 and 1, we have $|2\phi - 1| = 1$ which, from Proposition 1, implies that $AIM = |SIM| = PIN$. ■

[4] Lei and Wu (2005) are an exception as they allow strategic trading by informed investors.

Although the link between order imbalance and informed trading is not new [see Easley, Engle, O'Hara and Wu (2001) and Aktas et al. (2007)], our model allows us to deal with some important and realistic features of the process of informed trading. First, we show that the modeling and estimation of PIN requires an explicit recognition of the likelihood that the news arrival process and frequency are inherently different for different stocks. This, in turn, implies that the trading interval over which a PIN is estimated cannot be the same for all stocks.

The specification of the news arrival process is an integral part of the trading process in the framework proposed by Easley et al. (1996) and others. In their model, a news signal may arrive at the beginning of each trading interval and orders arrive throughout the remainder of the interval. However, there are important, yet unaddressed, issues in this framework. First, exactly how long should the trading interval last? Second, should the trading interval be of the same length for all stocks regardless of their characteristics? In fact, in the framework proposed by Easley et al. (1996), orders are always aggregated in a fixed trading interval, one day, for all stocks. This requires that news may arrive for all stocks at the beginning of each trading day. This seemingly innocuous assumption has important consequences that reflect poorly on the underlying news arrival process. The empirical observation of wide-spread difference in news coverage across companies in different industries attests to the need for firm-specific news arrival processes. Allowing news to arrive at most once a day appears to be too stringent for big companies not only because these firms frequently make public announcements and/or appear in news media, but also because they are covered by many stock analysts who release research reports from time to time. The aggregation of order flows on a daily basis for these stocks may consequently include periods during which news may have reversed, leading to the mitigation of the contrast between the informed and the uninformed traders. Therefore, the PIN measure based on daily data is likely to be underestimated for these stocks.

On the other hand, smaller stocks with little news coverage, potentially suffer from the opposite problem. It may very well be the case that it takes a couple of days, if not longer, for the current news to run its course before the arrival of a news reversal. Daily orders for such stocks would artificially inflate the contrast between the informed and the uninformed traders precisely due to the slow speeds of news arrival. As a result, the PIN measure can be overestimated in such instances. The important implication of the above discussion is that the optimal trading interval for estimating PIN may be different for stocks of different informational characteristics. Consequently, there may be systematic biases in inferences made in past studies about the existence and extent of information asymmetry across different stocks. For example, the well-documented pattern of high (or low) estimated PINs for less frequently (or frequently) traded stocks may be affected by the application of a simplistic, one-day-fits-all, news arrival process for all stocks. Similarly, the performance of the PIN-factor in explaining the cross-section of expected returns is likely to be influenced by systematic errors in the estimation of PIN.

In this paper, we analyze any patterns in estimated PINs over sampling intervals of different length. Using NYSE common stocks over the 1993-2006 period, we find lower estimated PINs associated with longer sampling intervals for the same stock over the same calendar quarter. This finding demonstrates the weakness of imposing an identical news structure over different stocks and applying a fixed aggregation interval for order flows.

Another implicit assumption in the framework by Easley, Kiefer, O'Hara and Paperman (1996) is that the uninformed traders submit orders in a balanced fashion (buy orders with 50% probability) whereas the informed traders solely drive the order imbalances on average. A natural inference is that high order imbalances make a stronger contrast between informed and uninformed traders, and thus reflect higher information asymmetry. Therefore, order imbalances (which are easily measurable) can be a legitimate proxy for information asymmetry just as the PIN measure (which can be obtained after costly estimations) was designed to be.

Following Proposition 2, and again using NYSE stocks over the 1993-2006 period, we estimate both PIN and AIM and demonstrate the relationship between them for different stocks and over different sampling intervals.

More importantly, we go one step further and provide a practical "rule of thumb" for choosing the different optimal trading intervals for estimating the PIN and AIM of different types of stocks. Although order imbalances and PIN will be linearly related, the exact equivalence between them will obtain only if there is no news reversals during the trading interval under consideration (see Propositions 1 and 2). We use this equivalence to arrive at an optimal trading interval for the estimation of PIN and, equivalently, AIM for each stock. We show that there is a vast variation in the length of this trading interval across different NYSE stocks. This implies that PIN estimates in past papers may suffer from systematic biases that may have led to erroneous time series and cross-sectional inferences.

Finally, we shed light on whether PIN or AIM reflects information asymmetry or liquidity. There are several papers that have successfully employed PIN as a measure of information asymmetry [see, for example, Easley et al. (1998), Easley, O'Hara and Saar (2001), Brown et al (2004), and Vega (2006)]. There is however a growing literature that questions the interpretation of PIN as a measure of information asymmetry [see, for example, Spiegel and Wang (2005), Mohanram and Rajgopal (2006), Atkas et al. (2007), and Duarte and Young (2007)]. Some of these papers suggest that PIN actually measures illiquidity and not asymmetric information. Similarly, a separate literature [see Chordia, Roll, and Subrahmanyam (2001, 2002, 2005) and Chordia and Subrahmanyam (2004)] has argued that order imbalance is also a liquidity measure. The disagreement in the literature about whether PIN or AIM reflect information asymmetry or liquidity is understandable because (a) both phenomena are very nebulous and consequently difficult to estimate, and (b) they invariably occur simultaneously.

We therefore conduct a novel type of test to gauge whether PIN or AIM indeed reflect changes in liquidity. The innovativeness of our test lies in using an exogenous event that should lead to changes in liquidity, but without any simultaneous variation in information asymmetry. Specifically, we examine the change in order imbalances around five national security events such as the attack on the World Trade Center in 2001. It is plausible that during these (daily) events, there could be a potential drop in stock liquidity, but little change in the extent of information asymmetry. We again use NYSE stocks over the 1993-2006 period, but exclude oil stocks, defense related stocks, and all stocks with earnings announcements close to these events in order to control for possible confounding effects. We find no statistical or economic changes in order imbalances over short intervals immediately surrounding these events, while two traditional measures of liquidity (quoted spread and effective spread) experience sharp economically and statistically significant declines. Although our experiment does not conclusively establish PIN and AIM as measures of asymmetric information, we provide a clean and strong test against the interpretation that they reflect liquidity.

II. Empirical Evidence

A. Data Construction

We use detailed trades and quotes information for stock transactions from the New York Stock Exchange (NYSE) Trade and Quote (TAQ) database for the period January 1993 to December 2006 and focus on common stocks that have been listed on the NYSE at least once during this period. We take great care in identifying the stocks properly because the stock ticker symbols can change over time and different CUSIPs can correspond to the same stock at different times. Following Chordia et al. (2001), we only use the primary market (NYSE) quotes because the auto-quotes are not filtered in TAQ. We retain quotes within the regular trading block, while purging those with non-positive bid or ask prices, negative bid or ask sizes, missing time stamps, or bid prices higher than ask prices. We also remove trades that are out of sequence, recorded before the open or after the close time, have special settlement conditions, or have missing trade size or time stamp. We use the Lee and Ready (1991) procedure to determine the buyer-initiated or seller-initiated nature of each trade, while imposing a five-second delay rule for matching trades with quotes prior to 1999. As recommended by Chordia et al. (2005), the five-second delay rule is revoked for matching trades with quotes starting in 1999. Finally, we keep only those trades that take place on the NYSE because the microstructure framework in Easley et al. (1996) applies to a specialist market.

The securities in our sample are sorted into deciles based on their trading intensity. Specifically, stocks are ranked each month based on average daily number of trades, and the monthly ranks are then assigned to all trading days in the month. We sort stocks into one of ten volume-deciles according to the average daily rank in each calendar quarter.

We begin with aggregating trades into 30-minute intervals. Starting at 9:30 a.m. each day, we count the number of buyer- and seller-initiated trades that occur in each half-hour interval until the market closes at 4:00 p.m. We also use five other aggregation intervals, leading to order flow data over a total of six sampling intervals: 30-minute, one-hour, half-day, one-day, two-day, and three-day. For each firm and calendar quarter, we separately estimate PINs using data constructed from all six sampling intervals.[5] Consequently, there are many more data points used to estimate PIN for the 30-minute interval than the 3-day interval.

We estimate PINs using the NLMIXED procedure in SAS. After some experimentation, we find that the quasi-Newton optimization method with BFGS updating provides the most stable parameter estimation in our setting. In about 3.5% of cases, the optimization routine was unable to find a global maximum of the likelihood function and/or generated problematic estimates for one or more inputs that were needed for the calculation of PIN. We remove these cases from our analysis.

B. PIN Estimates

The statistics of the estimated PINs are reported in Table I. Overall we obtain the most estimates of PIN at the daily sampling interval (84,144 firm-quarters) and the least estimates at the half-hour interval (80,426 firm-quarters). Nevertheless, the reported statistics are based on large samples for all sampling intervals.

Our empirical work highlights the difficulty of estimating PIN. In our sample of NYSE common stocks, we find that the non-convergence rates for the estimation of PIN over the 30-minute, one-hour, half-day, one-day, two-day and three-day intervals are 6.4%, 3.2%, 2.0%, 1.7%, 2.6% and 5.1%, respectively. When we split the stocks into ten volume-deciles within each calendar quarter, we find that the PIN estimation is more successful (in terms of convergence) among the intermediate volume-deciles and less successful at the extreme volume-deciles. The PIN estimation is more difficult for low-volume stocks because the lack of trades in these stocks often resulted in severely unbalanced orders (zero buy or sell orders) over a given time interval, producing numerical issues for the sample likelihood. The PIN estimation

[5] Details for the likelihood function and the trading process can be found in Easley et al. (1996).

is also difficult for the most-frequently traded stocks because the extreme number of trades causes numerical overflows in computing the factorials that are part of the likelihood function. A short sampling interval such as half an hour also compounds the problem because it becomes highly likely to have zero buy/sell orders over a very short period of time. A long sampling interval such as three days suffers little from the perspective of lacking orders on one side of the trades. Nevertheless, it contributes to problems of convergence along a different dimension because there are less data points available for each calendar quarter.

Two distinctive patterns emerge from a perusal of the estimates in Table I, Panel A, which contains the cross-sectional and time-series averages of estimated PINs over different combinations of volume-deciles and sampling intervals. The estimated PINs decline both in the length of sampling interval and the volume-decile. And, as we argue below, the two patterns, taken together, have some important, and hitherto undiscovered, implications for the estimation of the extent of asymmetric information in securities markets.

The first pattern is the rather striking decline in the estimated PINs with an increase in the sampling interval. This is an important pattern because the structural model underlying the estimation of PIN is silent as to the appropriate sampling interval for estimating PIN. Researchers have typically used one-day intervals, but there is no particular reason to believe that a 24-hour interval is more realistic than, say, half-day intervals or two-day intervals. As the results in Panel A of Table I make clear, the choice of sampling interval has a significant impact on the estimated PINs, a finding inconsistent with the theoretical prediction that the ratio of speeds of order arrival should be time invariant. Particularly, the mean estimated PINs almost always strictly decline in the length of the sampling intervals between one hour and three days. This pattern holds for stocks in each volume-decile.

Since this pattern of means could be affected by the fact that we are measuring mean estimates that are not strictly comparable because the firms underlying the different mean estimates may change across different sampling frequencies. We therefore also compare the estimated PINs for the same stock over increasing sampling intervals. Specifically, for each firm and in each calendar quarter, we compute the incremental PIN when the length of the sampling interval increases. Under each combination of volume-decile and sampling interval, we take the cross-sectional average of the firm-matched changes in PIN and report the time-series averages in Panel B. To examine whether these changes in estimated PINs deviate from zero, we report the Fama-MacBeth t-statistics in Panel C. Once again, we observe the pattern of declining PINs over increasing sampling intervals between one hour and three days, and such incremental changes in estimated PINs are statistically significant at the 1% level.[6]

There is an intuitive explanation for the systematic decline in the estimated PINs over the increasing length of the sampling intervals. The passage of time allows accumulation of order flows which, in turn, makes it very likely to reduce the order imbalance because the number of buyer-initiated trades will increasingly offset the number of seller-initiated trades with an increase in the sampling interval. One of the basic premises of the framework in Easley et al. (1996) is that on average only informed traders cause order imbalance, so it is no surprise that lower order imbalance over a longer time interval implies a lower level of estimated PIN.

The second pattern apparent in Table I, Panel A, is that the PIN estimates decline monotonically as the volume-decile increases, that is, high-volume stocks have lower PINs than low-volume stocks. We highlight the data for the one-day estimates in Panel A because past studies invariable use a 24-hour estimation interval for all stocks. The strictly monotonic negative relation with the volume-decile is consistent with the finding of Easley et al. (1996) in that they too document that high volume stocks have smaller PINs. These authors interpret this pattern as compelling empirical evidence that there exists large information asymmetry among thinly traded stocks.

Our first pattern of declining PIN estimates over the increasing sampling intervals, however, is related to the second pattern of the monotonic decline with volume-decile. It is noteworthy to see that there is a dramatic difference between the PIN estimates of the high- and low-volume deciles for the one-day and longer sampling intervals. For example, the one-day estimates for the lowest and highest volume-deciles are 0.2621 and 0.1035, respectively. Conversely, however, the difference drops dramatically at the high-frequency intervals. For

[6] One exception to this pattern is that for the lowest volume-decile the difference between the estimated PINs over half-day and one-hour intervals is positive. Another exception is that the mean estimated PINs increase, instead of declining, from the half-hour interval to the one-hour interval. The above deviations may have been caused by the large non-convergence rates of the PIN estimations over these two sampling intervals and for stocks in the lowest volume-decile.

example, the half-hour estimates for the lowest and highest volume-deciles are 0.2415 and 0.2046, respectively. This finding, if taken at face value, would counter-intuitively suggest there is relatively more asymmetric information at longer-sampling intervals. However, the first pattern uncovered by us suggests that the optimal sampling interval for estimating PIN should not be the same for all stocks. As we demonstrate below, the optimal sampling interval for stocks in the lowest volume-decile is likely to be significantly longer than the interval for stocks in the highest volume-decile because the news arrival frequency is likely to be different for the two types of stocks.

We next turn to the calculation of absolute order imbalances, AIM, for all stocks in our sample, their relationship with the PIN estimates and, most importantly, the determination of the optimal sampling intervals for estimating the degree of asymmetric information of stocks of different characteristics.

C. Order Imbalances

Having established and provided an intuitive account for the patterns in estimated PINs, we next examine whether or not the patterns in absolute (percentage) order imbalances (AIM) corroborate the intuition above. This also constitutes our first step to empirically verify the theoretical relation between PIN and AIM established in Section I. Table II, Panel A, contains the cross-sectional and time-series average estimates of AIM for all the stocks sorted into the same ten volume-deciles as Table I.

The two patterns uncovered for the PIN estimates are even stronger in the case of the AIM estimates. Without a single exception, the means of the AIMs decline strictly with the length of the sampling interval. As we increase the length of the sampling interval, the mean incremental changes in AIMs are all negative and statistically significant at the 1% level, as shown in Panels B and C of Table II. Similarly, there is a declining pattern in the AIMs with an increase in the volume-decile. This evidence is again consistent with the findings of Easley et al. (1996) that low-volume stocks are subject to a much larger extent of information asymmetry.

The close similarity in the patterns for the estimated PINs and AIMs is evidence in favor of a direct relationship between these two measures. As shown in Proposition 1 of Section I, a linear relationship between SIM and PIN should exist in the data. In addition, if there is no news reversal during the specific sampling interval under consideration, the two measures should be (theoretically) identical (see Proposition 2). We now provide more direct tests of the relationship between PIN and AIM, keeping in mind that highly-traded stocks are much more likely to receive frequent information than stocks that are traded infrequently. Before evaluating their relationship over different sampling intervals, we need to recognize however that both PIN and AIM are estimated with imprecision. Moreover, given the simplicity of AIM, and the corresponding complexity of estimating PIN, estimates of the latter are likely to be much more noisy especially for low-volume stocks and at high-frequency intervals.[7]

We calculate the firm-matched mean differences between the AIM and PIN for stocks in each of the ten volume-deciles during each sampling interval. Specifically, in each firm-quarter, we compute the difference between AIM and PIN corresponding to each of the six sampling intervals and report the cross-sectional and time-series average of such differences in Table III, Panel A. The deviations from the equivalence between AIM and PIN are positive for all but one reported combinations of volume-deciles and sampling intervals, and remain highly statistically significant (see Fama-MacBeth t-statistics in Table III, Panel B). Since we have a really large number of observations underlying each average, the statistical significance of the average differences between AIM and PIN (see Table III, Panel B) is not surprising. We now concentrate on the economic magnitude of the differences between the two estimates. In this context, it is important to recall that, although AIM and PIN will be related, the exact equivalence between PIN and AIM will obtain only during trading intervals that experience no news reversal [see Propositions 1 and 2].

One interesting pattern emerges from Table III, Panel A. The distances between AIM and PIN are monotonically declining in the trading volume and the length of sampling interval.

[7]The observed distances between AIM and PIN will also be related to the downward bias in estimated PINs illustrated by Boehmer et al. (2007). They show that errors in trade type classification tend to attenuate the observed magnitude of order imbalance relative to the true level and consequently lead to underestimated PINs. Moreover, Boehmer et al. (2007) document a larger downward bias for PINs among thinly traded stocks but caution that frequently traded stocks have a higher misclassification rate.

The smallest deviation is recorded at the level of -0.0015 for stocks in the highest trading volume-decile when the trade flows are aggregated every hour. Note that this is also one of the two only instances in which the deviation is not statistically different from zero. The natural implication is that for the most frequently traded stocks, it seems sufficient to aggregate the trade flows on an hourly basis and compute the absolute (percentage) order imbalance instead of estimating PINs.

It is intuitive that a natural trading interval can be very long for thinly traded stocks as the reversal of news requires passage of time, but that the optimal trading interval could be very short for frequently traded stocks. Therefore, the evidence in Table III again suggests that it does not seem reasonable to follow the conventional one-day-fits-all approach of using daily trade flows to estimate PINs for stocks of different characteristics. Likewise, a direct comparison of AIMs across different stocks may not be meaningful if the trade flows are all measured at the identical frequency.

D. The Optimal Sampling Interval

The direct test of the distance between PIN and AIM actually provides some empirical guidance as to the length of the sampling interval we should use for stocks with different news arrival and, therefore, trading intensities. Although our goal is not to determine the precise sampling interval for each stock, we want to nevertheless unequivocally establish that it is inappropriate to use the same (typically daily) sampling interval for gauging the extent of information asymmetry for all stocks.

We use multiple approaches to test for the equivalence between AIM and PIN. Recognize however that we do not know the true values of the parameters under consideration; we can only use estimates of both PIN and AIM. Although the estimation errors are likely to be more severe for PIN relative to AIM, we cannot set the exact equivalence between the two measures as the criterion for the choice of the optimal sampling interval. Moreover, we cannot rely on the statistical significance of the differences either as we have a large sample of firm-observations.

The first approach for determining the optimal sampling intervals for gauging the extent of information asymmetry for different stocks uses the mean differences between AIM and PIN tabulated in Panel A of Table III. Let us suppose that researchers set out a target distance between AIM and PIN of 0.05 as being an "acceptable" difference for a trading interval without a news reversal. Then the estimates in Table 3, Panel A suggest aggregating trade flows over an interval of half-hour to one-hour for the most frequently traded NYSE stocks. This benchmark has the appeal of also suggesting that the practice of aggregating daily order flows is appropriate for the average stock. Note that the one-day interval is appropriate for the fifth volume-decile. This, however, does not imply that the one-day interval is appropriate for all stocks. For stocks in the third volume-decile, for example, the suggested sampling interval is two days, and so on.

Given that both AIM and PIN are measured with error, we attempt to determine the equivalence between the two measures by estimating the following regression between AIM and PIN and evaluating the magnitude of the slope coefficient:

$$AIM = \alpha + \beta \cdot PIN + \varepsilon.$$

From Proposition 2, this slope coefficient, should be equal to 1.00 assuming no news reversal during the sampling interval deployed and no strategic behavior among informed traders. The cross-sectional and time-series averages of the estimated slope coefficients for each volume-decile during each of the six sampling intervals are presented in Table III, Panel C. Again, due to the large samples, we do not rely on statistical significance of the test that the slope coefficients are equal to 1.00; these tests are rejected in all but two of the 60 cases. For brevity, we omit reporting the Fama-MacBeth t-statistics. Instead, we again rely on economic departures from 1.00 as a rule of thumb.

The estimates of average betas across the different volume-deciles and the six sampling intervals present some very interesting patterns that strongly confirm that the optimal sampling intervals for different stocks should be quite different. For example, for stocks that trade very frequently and are therefore included in the highest volume-decile, the average betas decline systematically from a high of 1.18 to 0.82 for the half-hour to three-day intervals. This suggests that neither a very short nor a very long interval would be appropriate measurement interval for estimating the degree of asymmetric information for such stocks. Based on the beta estimates, an hourly or half-day interval would instead be appropriate.

Conversely, however, the behavior of the betas estimates for the lowest-decile portfolio containing infrequently traded stocks is exactly the opposite. The average betas start off at values close to zero for short intervals and increase steadily with the sampling interval to about 0.90 at the three-day interval. This evidence suggests that that both PIN and AIM estimates for infrequently traded stocks are largely noise at high-frequency intervals and a three-day (or even longer) interval may be optimal to measure the extent of asymmetric information for such stocks. Finally, a one-day sampling interval appears to be appropriate for only the "average" stock, one with a trading intensity that is neither too high nor too low.

To further confirm the conceptual basis of our practical "rules of thumb," we provide an alternative way of estimating the appropriate sampling interval for stocks in different volume-deciles. Since informed trades tend to have higher price impact than uninformed trades, we directly measure the average time it takes for a trade to exert a price impact of one basis point corresponding to stocks in each volume-decile. The time-to-price-impact variable is defined as the ratio of the time gap (in seconds) between every two consecutive trades relative to the change (in basis points) of trading prices over the same time span. For each firm during each trading day, we compute the weighted-average time-to-price-impact, with the volatility of per-trade returns (that is, the squared price changes in percentage) as weights. We then use estimates of the time-to-price-impact as an alternative way to determine the optimal sampling interval for measuring the degree of asymmetric information for different stocks.

The evidence in Table IV suggests that stocks with higher volume correspond to lower quoted spreads and shorter time-to-price-impact. And this pattern is strictly monotonic across the ten volume-deciles. It takes about 1.84 seconds on average to have a change of one basis point in trading price for the most frequently traded NYSE stocks. For the most thinly traded stocks, it takes 45.93 seconds on average. Now suppose that a price movement of 15%

represents a news reversal, then a natural trading interval for the highest volume-decile is about 46 minutes. By the same token, the natural trading interval for the lowest volume-decile is about 19 trading hours, and so on. Clearly, we can set a different cutoff rule for news reversal and utilize more stock-specific information to customize the determination of the appropriate length for the sampling interval of trade flows. Therefore, different researchers may prescribe varying lengths for the optimal sampling intervals.

We recognize that the choice of the optimal sampling intervals based on ad hoc cut offs of various measure is not precise. It does however constitute a simple, intuitively appealing, and very practical "rule of thumb." The important message is that we should aggregate trade flows over substantially different time intervals in order to estimate the degree of asymmetric information for stocks with different trading intensities. Perhaps equally importantly, it makes it possible to bypass the laborious process of estimating PINs; as AIM is much simpler to calculate and manipulate. This, in turn, has very important empirical implications because practitioners and academics alike are increasingly interested in the change of information content over rather short calendar time intervals, for which the conventional approach prescribed by Easley et al. (1996) does not work given the minimum requirement of 60 data points for the estimation of PINs. As mentioned earlier, aggregating hourly or half-day trade flows for the most frequently traded NYSE stocks should work well and it simultaneously addresses the non-convergence issue associated with estimating PINs for this particular group of stocks.

III. Information Asymmetry or Liquidity?

Given the theoretical link between the absolute (percentage) order imbalance (AIM) and the estimated probability of informed trading (PIN), it remains an empirical question whether they measure asymmetric information or liquidity. Several papers have successfully employed PIN as a measure of information asymmetry [see, for example, Easley et al. (1998), Easley, O'Hara and Saar (2001), Brown et al. (2004), and Vega (2006)]. There is however a growing literature that questions the interpretation of PIN as a measure of information asymmetry [see, for example, Spiegel and Wang (2005), Mohanram and Rajgopal (2006), Atkas et al.

(2007), and Duarte and Young (2007)]. Some of these papers suggest that PIN actually measures liquidity and not asymmetric information. Similarly, a separate literature [see Chordia, Roll, and Subrahmanyam (2001, 2002, 2005) and Chordia and Subrahmanyam (2004)] has argued that order imbalance is also a liquidity measure. The disagreement in the literature about whether PIN or AIM reflect information asymmetry or liquidity is understandable because (a) both phenomena are very nebulous and difficult to estimate, and (b) they invariably occur simultaneously.

We provide a new type of test to gauge whether PIN or AIM indeed reflect changes in liquidity. The crucial advantage of our test lies in the existence of exogenous events that should lead to predictable changes in liquidity, but plausibly little variation in information asymmetry. That is, we focus on environments during which it is plausible to have a minimal role for information asymmetry and a large role for liquidity effects. Specifically, we investigate the stock trading environments around a set of five major national security events as defined by Dow Jones Inc.[8] A national security event poses increased economic uncertainty and should make it harder for market participants to identify the true fundamental value of assets. Therefore, liquidity should decrease. Conversely, however, since a national security event presumably comes as a surprise to all market participants, it should not alter the extent of information asymmetry. In other words, during a national security event we should expect a trading environment with reduced liquidity and virtually no change in information asymmetry.

In this specific context, we examine the similarity, or the lack thereof, in the behavior of AIM relative to traditional measures of liquidity. If estimates of AIM move in tandem with traditional measures of liquidity, then it can be interpreted as strong evidence that AIM reflects liquidity. If, on the other hand, the AIM measure shows little correlation with the traditional liquidity measures, then it casts doubt on the liquidity interpretation.

We first identify national security event(s) during which measures of stock spreads (as a conventional proxy for stock liquidity) have increased dramatically, and then examine whether the AIM measure increases in a statistically significant manner to reflect the reduction in

[8] For details, see http://www.djindexes.com/mdsidx/index.cfm?event=showavgevents.

liquidity. For an event on trading day t, the abnormal change for each stock is defined as the difference between the variable measured on day t and its average value in the preceding 65 trading days excluding the immediately preceding five trading days. Two events, "Terrorist Attack" and "America Strikes Back", took place on a day when the market was closed, so we anchor the events on the next trading day immediately thereafter.

Not all common stocks on the NYSE are eligible for this exercise. Since the stocks in the security and defense sector may experience different trading patterns around these five national security events, we exclude 39 stocks from this analysis.[9] Also, given the prominent link between geopolitical concerns and petroleum price fluctuations, we remove 292 NYSE common stocks that are in the oil and gas related industries.[10] We also follow the standard practice in the microstructure literature and remove the high-priced Berkshire Hathaway stock from our sample. Finally, we remove all NYSE common stocks that have a quarterly earnings announcement close to the event date (plus or minus seven calendar days) so as to mitigate any potential confounding effects. We also sort stocks in the final sample into one of ten volume-deciles based on the daily average total trades in the month of each national security event.

We employ two conventional measures of liquidity: the quoted spread and the effective spread, the latter being defined as the difference between the actual trade price and the mid-point of the corresponding quote (both as percentages of the mid-point of the quote). The weights for the weighted average quoted spreads are the time each quote lasts until the next quote revision within the same trading day. On the other hand, arithmetic averages for each stock are computed for the daily effective spread. In the calculation of both measures of spread, we use only those trades and quotes that took place on the NYSE. Bessembinder (2003) provides a detailed analysis of different measures of spreads across different trading platforms, and we follow that study in constructing the spreads.

[9] We rely on Dun and Bradstreet's industry classification to identify 53 companies in the defense and security related industries. Among them 39 companies are listed on the NYSE and thus purged from our sample for this analysis.

[10] According to the CRSP database, there are 252 NYSE common stocks with share code 10 or 11 and 3-digit SIC code among 131, 132, 138, 291, 299, 492 or 493 for at least once in the 14-year period between 1993 and 2006.

Presumably, not all national security events have a similar impact on the trading activities of market participants. Those with a broader impact should register more pronounced drops in stock liquidity. Based on conventional measures of stock liquidity, the cross-sectional average abnormal changes reported in Table V suggest that three national security events are sub-optimal for our purpose. During the "World Trade Center Bombing," "Oklahoma Bombing," and "Operation Iraqi Freedom" events, the quoted spreads either increased by a statistically insignificant amount or actually declined. Moreover, there is a lack of any persistent pattern in the abnormal changes in quoted spreads across the ten volume-deciles. These three events also fail to witness a persistent drop in stock liquidity measured by the effective spread.

In sharp contrast, the "Terrorist Attack" and "America Strikes Back" stand out as two events with large decline in stock liquidity. On the trading day immediately following the "Terrorist Attack," both the quoted and effective spreads increased substantially for stocks in each of the ten volume-deciles. The abnormal changes are economically sizeable and statistically significant at the 1% level. Moreover, the abnormal increases in quoted and effective spreads decline monotonically as we move to stocks in higher volume-decile. Stocks in the lowest volume-decile (the most illiquid group) experience the most dramatic increase in percentage spreads, 140 basis points in quoted spread and 41 basis points in effective spread. Stocks in the highest volume-decile (the most liquid group) experience the smallest increase in percentage spreads, 21 basis points in quoted spread and 3 basis points in effective spread.

These magnitudes of changes in spreads may seem small at first glance, but this partially reflects the fact that spreads are small even during the benchmark period. For example, for the lowest volume-decile the ratio of the cross-sectional average spread on the first trading day after the "Terrorist Attack" relative to the average spread in the preceding three-month period is 1.5870 for quoted spread and 1.4482 for effective spread. In other words, the quoted and effective spreads increased by 58.70 percent and 44.82 percent, respectively. Therefore, the "Terrorist Attack" event was associated with a significant drop in liquidity.

During this national security event, however, the abnormal change in AIM is negative for all volume-deciles except the highest one. Although the most actively traded group of stocks witnessed a modest increase in AIM, the change is statistically insignificant. The widespread drop in AIM for all other volume-deciles is inconsistent with the interpretation of AIM as a liquidity measure.

On the trading day immediately after "America Strikes Back," there was also an abnormal increase in quoted spreads and effective spreads. Much like the "Terrorist Attack" event, this event was associated with abnormal increases in quoted and effective spreads that are both statistically significant (at the 1% level) and economically sizeable. The evidence on the cross-sectional average abnormal change in AIM again does not support a liquidity interpretation. While the increase in AIM for the second volume-decile is significant at the 5% level, changes in order imbalance experienced by stocks in all other nine volume-deciles are statistically indistinguishable from zero.

Based on the empirical design surrounding five national security events, we present strong evidence against interpreting the absolute order imbalance, and therefore PIN in light of their close relationship, as a measure of liquidity. There is a lack of any persistent increase in order imbalance following two specific exogenous events during which stocks experienced widespread and significant drops in liquidity. Also consistent with our conjecture that the national security events should have minimal impact on the extent of information asymmetry, we find that the changes in absolute order imbalance are not statistically significantly different from zero.

IV. Conclusion

In this study, we model and measure the existence of informed trading. Specifically, we investigate the properties of the widely used measure of informed trading, PIN, developed by Easley and O'Hara, and establish three important features of informed trading. First, the existence of informed trading, and therefore PIN, should be measured over different trading intervals for stocks of different characteristics. Second, we establish a useful relationship between PIN and the absolute percentage order imbalance (AIM). The latter is not only easier to measure, but can also be calculated over short horizons. Most importantly, the empirical equivalence between estimated PIN and AIM of a stock provides a guide to the estimation interval that should be used for that particular stock. Finally, an investigation around exogenous national security events reveals strong evidence against interpreting PIN and order imbalance as a liquidity measure.

V. Appendix

Proof of Proposition 1. Use the shorthand p for PIN and treat p as a known variable. Solve for $\{\phi, \iota, \nu\}$ from the following system of equations

$$\begin{aligned} B &= \phi\iota + \frac{1}{2}\nu, \\ S &= (1-\phi)\iota + \frac{1}{2}\nu, \\ p &= \frac{\iota}{\iota + \nu}. \end{aligned}$$

The solution turns out to be

$$\begin{aligned} \iota &= (B+S)p, \\ \nu &= (B+S)(1-p), \\ \phi &= \frac{B - \frac{1}{2}\nu}{\iota} = \frac{B - \frac{1}{2}(B+S)(1-p)}{(B+S)p}. \end{aligned}$$

The trading intensities over the interval Δt can be computed as

$$\begin{aligned} \frac{B}{B+S} &= \frac{1}{2} + p\phi - \frac{1}{2}p, \\ \frac{S}{B+S} &= \frac{1}{2} - p\phi + \frac{1}{2}p, \\ \frac{B-S}{B+S} &= p(2\phi - 1). \end{aligned}$$

■

References

Affleck-Graves, John, Carolyn M. Callahan, and Niranjan Chipalkatti, 2002, Earnings Predictability, Information Asymmetry, and Market Liquidity, *Journal of Accounting Research* 40, 561-583.

Akerlof, George A., 1970, The Market For "Lemons": Quality Uncertainty and the Market Mechanism, *Quarterly Journal of Economics* 84, 488-500.

Aktas, Nihat, Eric de Bodt, Fany Declerck, and Herve Van Oppens, 2007, The Pin Anomaly around M&A Announcements, *Journal of Financial Markets* 10, 169-191.

Amihud, Yakov, 2002, Illiquidity and Stock Returns: Cross-Section and Time-Series Effects, *Journal of Financial Markets* 5, 31-56.

Bessembinder, Hendrik, 2003, Issues in Assessing Trade Execution Costs, *Journal of Financial Markets* 6, 233-257.

Boehmer, Ekkehart, Joachim Grammig, and Erik Theissen, 2007, Estimating the Probability of Informed Trading–Does Trade Misclassification Matter?, *Journal of Financial Markets* 10, 26-47.

Brown, Stephen, Stephen A. Hillegeist, and Kin Lo, 2004, Conference Calls and Information Asymmetry, *Journal of Accounting and Economics* 37, 343-366.

Chen, Qi, Itay Goldstein, and Wei Jiang, 2007, Price Informativeness and Investment Sensitivity to Stock Price, *Review of Financial Studies* 20, 619-650.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, Commonality in Liquidity, *Journal of Financial Economics* 56, 3-28.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2001, Market Liquidity and Trading Activity, *Journal of Finance* 56, 501-530.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2002, Order Imbalance, Liquidity, and Market Returns, *Journal of Financial Economics* 65, 111-130.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2005, Evidence on the Speed of Convergence to Market Efficiency, *Journal of Financial Economics* 76, 271-292.

Chordia, Tarun, and Avanidhar Subrahmanyam, 2004, Order Imbalance and Individual Stock Returns: Theory and Evidence, *Journal of Financial Economics* 72, 485-518.

Duarte, Jefferson, Xi Han, Jarrad Harford, and Lance Young, 2008, Information Asymmetry, Information Dissemination and the Effect of Regulation Fd on the Cost of Capital, *Journal of Financial Economics* 87, 24-44.

Duarte, Jefferson, and Young Lance, 2007, Why Is Pin Priced?, *Journal of Financial Economics* forthcoming.

Easley, David, Robert F. Engle, Maureen O'Hara, and Liuren Wu, 2001, Time-Varying Arrival Rates of Informed and Uninformed Traders, *Cornell University Working Paper*.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is Information Risk a Determinant of Asset Returns?, *Journal of Finance* 57, 2185-2221.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2005, Factoring Information into Returns, *Cornell University Working Paper*.

Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman, 1996, Liquidity, Information, and Infrequently Traded Stocks, *Journal of Finance* 51, 1405-1436.

Easley, David, and Maureen O'Hara, 1992, Time and the Process of Security Price Adjustment, *Journal of Finance* 47, 577-605.

Easley, David, and Maureen O'hara, 2004, Information and the Cost of Capital, *Journal of Finance* 59, 1553-1583.

Easley, David, Maureen O'Hara, and Joseph Paperman, 1998, Financial Analysts and Information-Based Trade, *Journal of Financial Markets* 1, 175-201.

Easley, David, Maureen O'Hara, and Gideon Saar, 2001, How Stock Splits Affect Trading: A Microstructure Approach, *Journal of Financial and Quantitative Analysis* 36, 25-51.

Krinsky, Itzhak, and Jason Lee, 1996, Earnings Announcements and the Components of the Bid-Ask Spread, *Journal of Finance* 51, 1523-1535.

Lee, Charles M. C., Belinda Mucklow, and Mark J. Ready, 1993, Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis, *Review of Financial Studies* 6, 345-374.

Lee, Charles M. C., and Mark J. Ready, 1991, Inferring Trade Direction from Intraday Data, *Journal of Finance* 46, 733-746.

Lei, Qin, and Guojun Wu, 2005, Time-Varying Informed and Uninformed Trading Activities, *Journal of Financial Markets* 8, 153-181.

Mohanram, Partha, and Shiva Rajgopal, 2006, Is Information Risk (Pin) Priced?, *Columbia University Working Paper.*

Rothschild, Michael, and Joseph Stiglitz, 1976, Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, *Quarterly Journal of Economics* 90, 629-649.

Spence, Michael, 1974, Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution, *Journal of Economic Theory* 7, 296-332.

Spiegel, Matthew, and Xiaotong Wang, 2005, Cross-Sectional Variation in Stock Returns: Liquidity and Idiosyncratic Risk, *Yale University Working Paper.*

Vega, Clara, 2006, Stock Price Reaction to Public and Private Information, *Journal of Financial Economics* 82, 103-133.

Venkatesh, P. C., and R. Chiang, 1986, Information Asymmetry and the Dealer's Bid-Ask Spread: A Case Study of Earnings and Dividend Announcements, *Journal of Finance* 41, 1089-1102.

Yan, Yuxing, and Shaojun Zhang, 2006, An Improved Estimation Method and Empirical Properties of Pin, *SSRN Working Papers* 890486.

Table I. Estimated PINs

This table presents the results of estimated probability of informed trading (PINs) for all NYSE common stocks between January 1993 and December 2006. Stocks are sorted into one of ten volume deciles in each calendar quarter. We use the Fama-MacBeth approach to compute the cross-sectional average PIN for each quarter and report the time-series averages in Panel A. As the length of the sampling interval increases, the incremental changes in PINs are computed for a given firm and calendar quarter. Again, we compute the cross-sectional average changes in each quarter and report the time-series average in Panel B. To examine whether these changes deviate from zero, we report the corresponding Fama-MacBeth t-statistics in Panel C.

Volume Decile	Half Hour	Hour	Half Day	1 Day	2 Day	3 Day
Panel A. Mean of Estimated PINs						
Lowest	0.2415	0.2654	0.2757	**0.2621**	0.2420	0.2316
2	0.2387	0.2579	0.2351	**0.2093**	0.1814	0.1724
3	0.2427	0.2591	0.2146	**0.1882**	0.1634	0.1558
4	0.2521	0.2646	0.2002	**0.1757**	0.1530	0.1460
5	0.2521	0.2654	0.1854	**0.1619**	0.1411	0.1358
6	0.2522	0.2660	0.1742	**0.1517**	0.1328	0.1289
7	0.2530	0.2642	0.1638	**0.1415**	0.1242	0.1202
8	0.2490	0.2584	0.1518	**0.1310**	0.1154	0.1119
9	0.2365	0.2473	0.1384	**0.1193**	0.1071	0.1043
Highest	0.2046	0.2229	0.1175	**0.1035**	0.0982	0.0975
Panel B. Mean Incremental Changes in Estimated PINs						
Lowest		0.0157	0.0055	-0.0146	-0.0212	-0.0102
2		0.0174	-0.0243	-0.0257	-0.0279	-0.0097
3		0.0156	-0.0453	-0.0265	-0.0248	-0.0078
4		0.0113	-0.0645	-0.0246	-0.0224	-0.0072
5		0.0124	-0.0800	-0.0234	-0.0208	-0.0054
6		0.0135	-0.0918	-0.0222	-0.0191	-0.0040
7		0.0113	-0.1005	-0.0222	-0.0173	-0.0043
8		0.0093	-0.1065	-0.0209	-0.0158	-0.0039
9		0.0107	-0.1089	-0.0190	-0.0125	-0.0032
Highest		0.0181	-0.1052	-0.0142	-0.0062	-0.0026
Panel C. t-statistic for Mean Changes in Estimated PINs						
Lowest		6.69	1.65	-14.56	-19.08	-12.30
2		9.38	-4.32	-21.99	-27.79	-10.96
3		8.78	-9.44	-25.77	-30.91	-7.29
4		4.65	-19.04	-23.42	-30.79	-6.04
5		6.01	-37.61	-20.31	-21.83	-5.25
6		8.32	-83.66	-19.03	-19.30	-3.85
7		5.48	-77.54	-16.81	-18.93	-4.40
8		3.74	-53.04	-15.75	-18.67	-3.46
9		3.95	-50.38	-16.93	-15.75	-3.14
Highest		6.82	-71.21	-14.71	-3.21	-1.88

Table II. Absolute Percentage Order Imbalances

This table presents the results of absolute percentage order imbalances (AIMs) for all NYSE common stocks January 1993 and December 2006. Stocks are sorted into one of ten volume deciles in each calendar quarter. We use the Fama-MacBeth approach to compute the cross-sectional average AIM for each quarter and report the time-series averages in Panel A. As the length of the sampling interval increases, the incremental changes in AIMs are computed for a given firm and calendar quarter. Again, we compute the cross-sectional average changes in each quarter and report the time-series average in Panel B. To examine whether these changes deviate from zero, we report the corresponding Fama-MacBeth t-statistics in Panel C.

Volume Decile	Half Hour	Hour	Half Day	1 Day	2 Day	3 Day
Panel A. Mean of AIMs						
Lowest	0.8387	0.7840	0.6025	0.4862	0.3799	0.3309
2	0.7052	0.6319	0.4149	0.3169	0.2445	0.2145
3	0.6335	0.5563	0.3450	0.2666	0.2099	0.1862
4	0.5741	0.4965	0.3002	0.2366	0.1893	0.1697
5	0.5189	0.4437	0.2640	0.2108	0.1707	0.1539
6	0.4692	0.3999	0.2380	0.1921	0.1574	0.1430
7	0.4209	0.3591	0.2137	0.1738	0.1433	0.1307
8	0.3732	0.3204	0.1916	0.1571	0.1304	0.1198
9	0.3248	0.2815	0.1695	0.1406	0.1185	0.1100
Highest	0.2512	0.2209	0.1368	0.1171	0.1026	0.0972
Panel B. Mean Incremental Changes in AIMs						
Lowest		-0.0547	-0.1815	-0.1163	-0.1063	-0.0490
2		-0.0733	-0.2170	-0.0980	-0.0724	-0.0299
3		-0.0772	-0.2113	-0.0784	-0.0568	-0.0236
4		-0.0777	-0.1963	-0.0636	-0.0472	-0.0197
5		-0.0752	-0.1797	-0.0532	-0.0401	-0.0167
6		-0.0693	-0.1620	-0.0459	-0.0347	-0.0144
7		-0.0618	-0.1454	-0.0399	-0.0305	-0.0126
8		-0.0528	-0.1288	-0.0345	-0.0267	-0.0106
9		-0.0433	-0.1120	-0.0289	-0.0221	-0.0085
Highest		-0.0304	-0.0841	-0.0197	-0.0145	-0.0054
Panel C. t-statistic for Mean Changes in AIMs						
Lowest		-43.99	-97.20	-35.35	-34.58	-31.47
2		-32.36	-28.88	-16.61	-19.38	-19.40
3		-21.06	-20.73	-14.81	-18.39	-17.40
4		-17.23	-17.91	-14.05	-17.27	-15.81
5		-15.18	-16.44	-13.91	-16.82	-15.19
6		-13.91	-15.55	-13.66	-16.44	-15.05
7		-12.94	-14.95	-13.51	-16.03	-14.67
8		-12.17	-14.23	-12.99	-15.87	-14.36
9		-11.98	-13.90	-12.77	-15.84	-13.57
Highest		-11.81	-13.98	-12.68	-15.61	-12.70

Table III. Differences in AIM and PIN

This table presents the results of two tests. First, we test whether or not the absolute percentage order imbalances (AIMs) are the same as the estimated probability of informed trading (PINs) for the same firm over the same calendar quarter. For a given firm and calendar quarter, we compute the average AIMs at each sampling frequency, from which we then subtract the estimated PINs. We use the Fama-MacBeth approach to compute the cross-sectional average differences for each quarter and report the time-series averages in Panel A. To examine whether these differences deviate from zero, we report the corresponding Fama-MacBeth t-statistics in Panel B. Second, we run a cross-sectional regression of AIM on PIN for each quarter, $AIM = \alpha + \beta \cdot PIN + \varepsilon$, and report in Panel C the time-series average slope coefficient $\widehat{\beta}$ for PIN.

Volume Decile	Half Hour	Hour	Half Day	1 Day	2 Day	3 Day
			Panel A. Mean of (AIM - PIN)			
Lowest	0.5900	0.5115	0.3219	0.2204	0.1356	0.0971
2	0.4659	0.3737	0.1799	0.1076	0.0631	0.0427
3	0.3905	0.2970	0.1304	0.0785	0.0464	0.0305
4	0.3218	0.2318	0.1000	0.0610	0.0363	0.0238
5	0.2667	0.1783	0.0785	0.0489	0.0296	0.0181
6	0.2169	0.1339	0.0639	0.0403	0.0246	0.0142
7	0.1679	0.0949	0.0499	0.0322	0.0190	0.0105
8	0.1242	0.0620	0.0397	0.0260	0.0151	0.0083
9	0.0882	0.0342	0.0312	0.0213	0.0117	0.0064
Highest	0.0468	-0.0015	0.0193	0.0139	0.0058	0.0017
			Panel B. t-statistic for Mean of (AIM - PIN)			
Lowest	36.52	31.01	23.04	21.31	19.95	18.53
2	16.42	14.19	12.46	12.64	12.54	10.84
3	13.74	11.77	11.65	12.00	11.65	10.26
4	12.64	10.42	11.51	11.86	12.47	10.72
5	11.81	9.36	11.63	12.02	13.02	10.48
6	11.14	8.41	11.82	12.36	13.95	9.38
7	10.49	7.19	11.24	11.64	12.15	8.46
8	9.69	5.70	10.82	11.89	11.48	7.18
9	8.90	3.85	10.75	12.10	10.03	4.61
Highest	7.05	-0.23	10.32	10.52	3.01	0.83
			Panel C. Estimated Slope Coefficient for PIN			
Lowest	-0.0195	-0.0029	0.4598	0.7192	0.8587	0.9000
2	0.3400	0.4317	0.6119	0.6487	0.6785	0.7160
3	0.3893	0.4771	0.6549	0.6801	0.7137	0.7203
4	0.4749	0.5881	0.7048	0.7181	0.7348	0.7472
5	0.5481	0.6713	0.7246	0.7251	0.7346	0.7284
6	0.6121	0.6900	0.7299	0.7446	0.7463	0.7439
7	0.7181	0.7681	0.7708	0.7815	0.7749	0.7885
8	0.7825	0.8231	0.8043	0.8102	0.8053	0.8027
9	0.8770	0.8640	0.8389	0.8535	0.8350	0.8323
Highest	1.1839	1.1197	0.9717	0.9749	0.9486	0.8215

Table IV. Time to Price Impact and Quoted Spread

This table presents the summary statistics for two variables using all NYSE common stocks between January 1993 and December 2006. The time to price impact variable is defined as the ratio of the time gap (in seconds) between every two consecutive trades to the change (in basis points) of trading prices over the same time span. For each firm on each trading day, we compute the weighted-average time to price impact, with the volatility of per trade returns (e.g., the squared price changes in percentage) as weights. The quoted spread is defined as the quoted bid-ask spread relative to the mid-quote (in basis points). We use the Fama-MacBeth approach to compute the cross-sectional averages of these two variables for each quarter, and report the time-series average as well as the Fama-Macbeth t-statistics.

Volume Decile	Time to Price Impact		Quoted Spread	
	Mean	t-statistic	Mean	t-statistic
Lowest	45.93	23.35	239.98	16.03
2	23.77	16.93	123.98	13.14
3	16.46	15.42	98.67	11.98
4	12.39	15.13	77.92	11.59
5	9.67	15.13	61.04	11.26
6	7.29	16.18	50.62	11.12
7	5.88	16.23	42.04	11.13
8	4.69	16.19	34.97	11.07
9	3.46	16.98	28.99	11.21
Highest	1.84	19.54	22.65	11.43

Table V. Abnormal Changes around National Security Events

This table presents the pattern of abnormal changes around five national security events that are defined by Dow Jones Inc. The variables of interest include the absolute percentage order imbalances, the quoted spreads and the effective spreads. For an event on trading day t, the abnormal change for each stock is defined as the difference between the variable measured on day t and the average value in the preceding 65 trading days excluding the immediately preceding five trading days. Two events, "Terrorist Attack" and "America Strikes Back", took place on a day when the market was closed, so we use the variables measured on the next trading day immediately thereafter instead. We exclude security, defense, oil, gas related stocks and stocks with quarterly earnings announcements close to the event date (plus or minus seven calendar days). The numbers reported in five panels, one for each event, are the cross-sectional average abnormal changes for all NYSE common stocks in ten volume deciles. The p-values for the cross-sectional averages are inside parentheses.

Volume Decile	Absolute Imbalance		Quoted Spread		Effective Spread	
Panel A. "World Trade Center Bombing" on Februay 26, 1993						
Lowest	0.080	(0.06)	-39.558	(0.04)	-16.033	(0.11)
2	0.048	(0.14)	15.890	(0.07)	3.956	(0.48)
3	0.079	(0.01)	2.689	(0.72)	1.108	(0.68)
4	0.018	(0.42)	-7.679	(0.34)	-2.024	(0.39)
5	0.042	(0.07)	11.260	(0.04)	3.034	(0.06)
6	-0.010	(0.62)	4.736	(0.08)	1.488	(0.08)
7	-0.001	(0.94)	1.036	(0.85)	-1.706	(0.51)
8	0.015	(0.42)	5.369	(0.16)	-0.584	(0.57)
9	-0.016	(0.21)	2.575	(0.47)	0.467	(0.42)
Highest	0.012	(0.27)	-0.049	(0.97)	0.508	(0.19)
Panel B. "Oklahoma Bombing" on April 19, 1995						
Lowest	0.118	(0.00)	-33.451	(0.15)	6.310	(0.52)
2	0.072	(0.08)	-7.933	(0.39)	-4.754	(0.21)
3	0.032	(0.33)	-1.987	(0.79)	1.178	(0.74)
4	0.041	(0.17)	-1.706	(0.74)	-0.726	(0.68)
5	-0.022	(0.42)	3.141	(0.47)	0.233	(0.91)
6	0.020	(0.51)	0.768	(0.92)	-0.339	(0.93)
7	0.028	(0.36)	-2.296	(0.42)	-1.827	(0.02)
8	-0.018	(0.50)	-7.141	(0.05)	-2.072	(0.05)
9	-0.010	(0.66)	-0.426	(0.85)	-0.771	(0.41)
Highest	-0.022	(0.16)	2.606	(0.08)	-0.207	(0.59)
Panel C. "Terrorist Attack" on September 11, 2001						
Lowest	-0.090	(0.01)	140.320	(0.00)	40.984	(0.00)
2	-0.054	(0.00)	86.129	(0.00)	32.962	(0.00)
3	-0.023	(0.12)	78.721	(0.00)	23.961	(0.00)
4	-0.014	(0.29)	45.297	(0.00)	15.035	(0.00)
5	-0.027	(0.01)	42.492	(0.00)	12.756	(0.00)
6	-0.020	(0.09)	25.223	(0.00)	6.649	(0.00)
7	-0.033	(0.00)	24.857	(0.00)	5.681	(0.00)
8	-0.017	(0.07)	25.144	(0.00)	3.758	(0.00)
9	-0.015	(0.11)	21.688	(0.00)	3.554	(0.00)
Highest	0.006	(0.40)	20.980	(0.00)	3.286	(0.00)

Volume Decile	Absolute Imbalance		Quoted Spread		Effective Spread	
	Panel D. "America Strikes Back" on October 7, 2001					
Lowest	0.043	(0.23)	72.277	(0.01)	34.107	(0.01)
2	0.056	(0.04)	48.987	(0.00)	21.657	(0.00)
3	0.017	(0.32)	31.644	(0.00)	11.069	(0.00)
4	0.024	(0.13)	20.405	(0.00)	5.964	(0.00)
5	-0.007	(0.57)	13.272	(0.00)	3.355	(0.00)
6	-0.021	(0.08)	5.032	(0.00)	2.019	(0.00)
7	-0.015	(0.13)	3.873	(0.00)	1.472	(0.00)
8	0.007	(0.46)	3.133	(0.00)	1.089	(0.00)
9	0.007	(0.48)	2.593	(0.00)	1.214	(0.00)
Highest	0.001	(0.87)	2.020	(0.00)	0.952	(0.00)
	Panel E. "Operation Iraqi Freedom" on March 19, 2003					
Lowest	-0.016	(0.58)	17.897	(0.11)	1.306	(0.78)
2	-0.004	(0.80)	1.934	(0.58)	1.121	(0.43)
3	0.013	(0.37)	6.618	(0.00)	2.120	(0.01)
4	0.021	(0.12)	1.699	(0.12)	0.335	(0.36)
5	0.019	(0.09)	0.251	(0.78)	0.545	(0.37)
6	0.007	(0.45)	-1.827	(0.00)	-0.603	(0.00)
7	-0.003	(0.74)	-0.368	(0.39)	0.004	(0.98)
8	0.014	(0.13)	-0.549	(0.37)	-0.083	(0.64)
9	0.039	(0.00)	-0.055	(0.95)	-0.038	(0.81)
Highest	0.040	(0.00)	-1.359	(0.00)	-0.235	(0.00)