Economics Bulletin

Volume 31, Issue 4

Intraday probability of informed trading

Zaichao Du
Southwestern University of Finance and Economics

Abstract

By extending Easley, Kiefer, O'Hara and Paperman's (1996) framework to an intraday model, I empirically estimate the intraday probability of informed trading (PIN) for the 30 stocks in DJIA index. I document a U-shaped PIN pattern over the time of a day, and the consequent test validates this finding.

I thank Professor Craig Holden and Dr. Konstantin Tyurin for their helpful comments. All the errors are my own. Research funded by Western and Border Areas Program, Humanities and Social Sciences, Ministry of Education of China, 11XJC790002.
Citation: Zaichao Du, (2011) "Intraday probability of informed trading", *Economics Bulletin*, Vol. 31 No. 4 pp. 3103-3112.
Contact: Zaichao Du - zadu@umail.iu.edu.
Submitted: August 23, 2011. **Published:** November 09, 2011.

1. Introduction

Informed trading is an important issue in market microstructure. It can partly explain the existence of the bid-ask spread. The core idea is that the market maker faces an adverse selection problem, since a customer agreeing to trade at the market maker's ask or bid price may be trading because he knows something that the market maker does not. In effect, then, the market maker must recoup the losses suffered in trades with the well informed by gains in trades with uninformed traders. These gains are achieved by setting a positive spread. This idea is one of the fundamental insights of the theoretical microstructure literature, see Kyle (1985) and Glosten and Milgrom (1985).

This paper links two important areas of the informed trading literature. The first area is PIN models. Easley, Kiefer, O'Hara and Paperman (1996), Easley, Kiefer and O'Hara (1997) developed a framework for analyzing the information in the trading process, in which a competitive market maker sets bid and ask prices based on her forecast of the composition of the traders (informed versus uninformed) and the probability of good or bad news. Their model started a whole series of market microstructure literatures on the probability of informed trading (PIN), such as Easley, O'Hara and Paperman (1998), Easley, Engle, O'Hara and Wu (2001), Easley, Hvidkjaer and O'Hara (2002), (2005) and so on. Using PIN models, these authors addressed and explained lots of practical issues, like the role of purchased order flow, how stock splits affect trading, whether information risk is a determinant of asset returns, the role of financial analysts and so on. We can see the PIN models are pretty powerful, but at the same time we may notice that all these PIN estimates and applications are based on daily intervals of time aggregation.

The second area concerns some intraday issues. For example, Bloomfield et al (2005) find that liquidity provision evolves during the trading day. They find experimentally that informed traders demand liquidity early in the trading session by submitting market orders but supply liquidity later by submitting limit orders. Chung et al (1999) document a U-shaped pattern in trading activity; Barclay and Hendershott (2003) document a U-shaped pattern in price discovery over the trading day with a much larger spike at the beginning of the day. This suggests that there is more informed trading early in the day. Thus, it is possible that information accumulates overnight when trading is more costly. For example, Greene and Watts (1996) find that two-thirds of sample earnings announcements take place after the stock exchanges close. This implies that many informed traders face an end-of-day deadline for exploiting their information. Anand et al (2005) find that informed limit orders are more profitable early in the day than later and that medium-sized institutional marketable orders (their proxy for informed market orders) account for more of the cumulative price movement in the morning than in the afternoon. The latter result suggests that informed investors are most aggressive in their use of market orders early in the day. [1] Most of these intraday models focus on the intraday patterns of bid-ask spread, trading volume, transaction cost and so on. Although some people did mention the possible pattern of informed trading over time of day, nobody actually estimated or tested intraday PINs.

[1] See also Ellul, Holden, Jain and Jennings (2006)

Connecting the PIN models and the intraday models together, my research questions are, does PIN change over time of day? If yes, how does it change? Can we find U-shaped pattern in probability of informed trading also?

To answer these questions, I extend Easley, Kiefer, O'Hara and Paperman's 1996 framework to an intraday model, and empirically estimate the model's parameters from a time series of trade data. These parameters are the "primitives" underlying the market-maker's learning and pricing problem, and from which we can estimate the probability of informed trading (PIN).

After estimating the intraday PINs for the 30 stocks in the Dow Jones Industrial Average Index, I found that PIN does change over time of day. Interestingly enough, the PIN does not change monotonically in a day, instead it starts high in the early time of day, and decreases till the middle of day, then rises up again near the end of a day, i.e. we found U-shaped pattern of intraday PINs also. Admati and Pfleiderer's (1988) explanation for the high PIN near the end of day is that some liquidity traders have discretion over when they trade, they prefer to trade when the market is "thick"-that is, when their trading has little effect on prices. This creates strong incentives for liquidity traders to trade together and for trading to be concentrated. When informed traders can also decide when to collect information and when to trade, they also want to trade when the market is thick. If many informed traders trade at the same time that liquidity traders concentrate their trading, then the terms of trade will reflect the increased level of informed trading as well.

Hong and Wang (2000) explained the possible U-shaped PIN pattern over the day as follows. During the closure, market prices cease to provide information to the uninformed investors. As the informed investors continue to receive new private information, the information asymmetry between the two classes of investors increases over the night. At the open, trading resumes. The opening price partially reveals the private information of the informed investors accumulated over the closure, causing the level of information asymmetry to drop discretely. As trading continues during the day, more private information is revealed through the prices, and the level of information asymmetry may continue to decrease. At the close, two additional forces come into play. On the one hand, the informed investors cut back their hedging positions, which also unveils their speculative positions and their private information. This tends to speed the decrease in information asymmetry at the close. On the other hand, the informed investors also cut back their speculative positions due to the high risk over the closure. The reduction in speculative trade makes the stock price less revealing about existing private information. This tends to increase the information asymmetry at the close. When the second force dominates the first one, there will be a high information asymmetry around the close also.

The rest of the paper is organized as follows. In section 2 we extend the Easley, Kiefer, O'Hara and Paperman (1996) theoretical microstructure framework to an intraday model. In section 3 we describe our data for the 30 stocks in the Dow Jones Industrial Average Index, and present our maximum likelihood estimation results. We also discuss the reliability of our estimates, and then investigate the interpretation and implications of our results. Section 4 provides a summary of our results, and concludes with a discussion of some directions for future research.

2. The Theoretical Model

The theoretical model used in this paper is based on Easley, Kiefer, O'Hara and Paperman's (1996) framework, where potential buyers and sellers trade a single asset with a market maker. The market maker is risk neutral and competitive, and quotes prices at which she will buy or sell the asset. Traders arrive at the market according to some Poisson processes, and trade at the quoted prices.

I extend this framework into an intraday model. Following Madhavan, Richardson and Roomans (1997) I divide one day into five trading periods: 9:30am to 10am, 10am to 11:30am, 11:30am to 2pm, 2pm to 3:30pm, and 3:30pm to 4pm. Prior to the beginning of any trading period, nature determines whether an information event relevant to the value of the asset will occur. Information events are independently distributed and occur with probability α_i, i =1, 2, 3, 4, 5. These events are good news with probability 1 - δ_i, or bad news with probability δ_i.

If there is an information event in period i, then those traders who have seen the new information will arrive at the market and trade according to a Poisson process with arrival rate μ_i. The informed traders will sell the asset if they have observed bad information, and they will buy it if they have observed good information. If no information event has occurred in period i, then only the uninformed buyers and uninformed sellers will trade in that period for liquidity reasons, and they each arrive at rate ε_i. The tree in Figure 1 depicts the trading process.

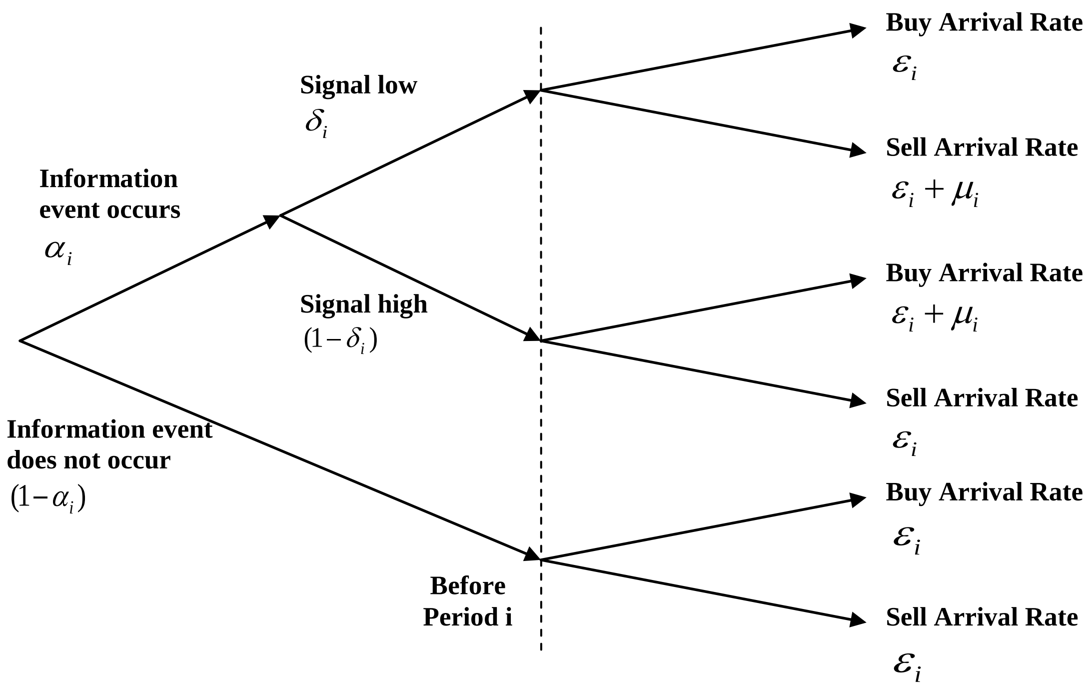


Figure 1, Tree diagram of the trading process.
This figure gives the structure of the trading process, where α_i is the probability of an information event, δ_i is the probability of a low signal, μ_i is the rate of informed trade arrival, and ε_i is the rate of uninformed buy and sell trade arrivals. Nodes to the left of the dotted line occur only once before period i.

As we can see, the probabilistic structure of our modified model is completely described by the parameters α_i, δ_i, μ_i, and ε_i. We are going to empirically estimate them from a time series of trade data. Given those estimates, we could calculate the PINs over the day.

As in Easley, Kiefer and O'Hara (1997), we can show that the total number of buys B_i, and sells S_i are sufficient statistics for the estimation purpose. After dropping some constant terms, our log likelihood function is given by

$$\sum_{d=1}^{D}\sum_{i=1}^{5}\{[-2\varepsilon_i + M_i \ln x_i + (B_i + S_i)\ln(\mu_i + \varepsilon_i)]$$
$$+\ln[\alpha_i(1-\delta_i)e^{-\mu_i}x^{S_i-M_i} + \alpha_i\delta_i e^{-\mu_i}x_i^{B_i-M_i} + (1-\alpha_i)x_i^{B_i+S_i-M_i}]\}, \tag{1}$$

where D is the total number of days, $M_i \equiv \min(B_i, S_i) + \max(B_i, S_i)/2$, $x_i \equiv \dfrac{\varepsilon_i}{\varepsilon_i + \mu_i}$.

We are going to get the estimates for the parameters α_i, δ_i, μ_i, and ε_i by maximizing the log likelihood function of equation (1). After that we can calculate PIN_i using the expression $PIN_i = \dfrac{\alpha_i\mu_i}{\alpha_i\mu_i + 2\varepsilon_i}$.

3. Data and Estimation

The estimation described in the last section requires trade outcome data for a specific stock over some sample period. For the analysis in this paper, we selected the 30 stocks in the Dow Jones Industrial Average Index to be our sample stocks. This selection was dictated both by convenience and by the relatively active trading found in those 30 stocks. This latter characteristic is important given that it is the information contained in trade data that is the focus of our work.

Trade data for the 30 stocks were taken from the NYSE TAQ database for the period September 1, 2006, to November 30, 2006. A 63 trading-day window was chosen to allow sufficient trade observations for our estimation procedure. The TAQ data provide a complete listing of quotes, depths, trades, and volume at each point in time for each traded security.

For our analysis, we require the number of buys and sells in each period for each stock-day in our sample. Since these are not immediately obtainable from the data, a number of transformations were needed to derive our data.

The most important transformation to the data involves the classification of buy and sell trades. Our model requires these to be identified, but the TAQ data record only transactions, not who initiated the trade. Here we use a technique developed by Lee and Ready (1991). Those authors propose defining trades above the midpoint of the prevailing bid-ask spread to be buys and trades below the midpoint of the prevailing spread to be sells. Trades at the midpoint are classified depending upon the price movement of the previous trade. Thus, a midpoint trade will be a sell if the midpoint moved down from the previous trade (a downtick) and will be a buy if the midpoint moved up. If there was no price movement then we move back to the prior price movement and use that as our benchmark. But according to Henker and Wang (2006), we use the National Best

Bid and Offer price (NBBO) 1 second ahead of trade as the prevailing quotes. Applying this modified algorithm to each transaction in our sample, we determined the numbers of buys and sells in each trading period for each stock-day. The resulting trade outcome data are given in Table 1, where I only reported the average across the 30 stocks. Because the durations of the five periods are different, the numbers are reported per half-hour, so that one can compare across periods.

Table 1. Trade Process Data

	9:30-10:00	10:00-11:30	11:30-14:00	14:00-15:30	15:30-16:00
No. Buys (per half hour)	1143	867	587	752	1209
No. Sells (per half hour)	1073	832	565	705	1097

Trade data for all the 30 stocks in the DJIA index for the period September 1, 2006 to November 30, 2006. The number of buys and sells is determined from transactions data using the modified Lee-Ready algorithm, and they are averaged across time series first, then cross-section.

As is apparent, the number of buys and sells varies across periods. Transactions take place more frequently at the beginning and end of the day, compared with the middle of the day, which is consistent with the U-shaped pattern in trading volumes over a day.[2]

From the above time series of average trade data, we can estimate the parameters α_i, δ_i, μ_i, and ε_i by running the MLE procedure in SAS. From those parameters we get the estimates for PIN_i using the expression $PIN_i = \frac{\alpha_i \mu_i}{\alpha_i \mu_i + 2\varepsilon_i}$. The results are in Table 2.

One may notice that we started with 30 stocks, but we only reported the results for 18 stocks here. What happened is the MLE procedure failed to converge in at least three periods for the other 12 stocks. The main reason for the failure is the overflow of the numerical derivatives, which is not surprising considering the complexity of our log likelihood function. Since the number of buys and sells is in thousands sometimes, the calculations of the derivatives involve some numbers to the thousandth power, which easily causes overflow.

But we still get the PIN estimates for 18 stocks, which is enough for our analysis. More importantly, those estimates are pretty reliable. Actually most of our estimates for α_i, δ_i, μ_i, and ε_i are significantly different from 0 at 5% level. Besides, the MLE procedure converges to the stationary parameter estimates very fast from a variety of starting values. Furthermore, the gradient of the likelihood function at our estimated parameters is less than 10^{-5} in absolute value, although our log likelihood function contains some numbers to the thousandth power, which suggests that our estimates are pretty accurate. So I'll continue the analysis with the results for the 18 stocks.

[2] Many people have documented this fact. For example Madhavan, Richardson, Roomans (1997)

Economics Bulletin, 2011, Vol. 31 No. 4 pp. 3103-3112

Table 2. PIN Estimates for Certain Stocks in DJIA

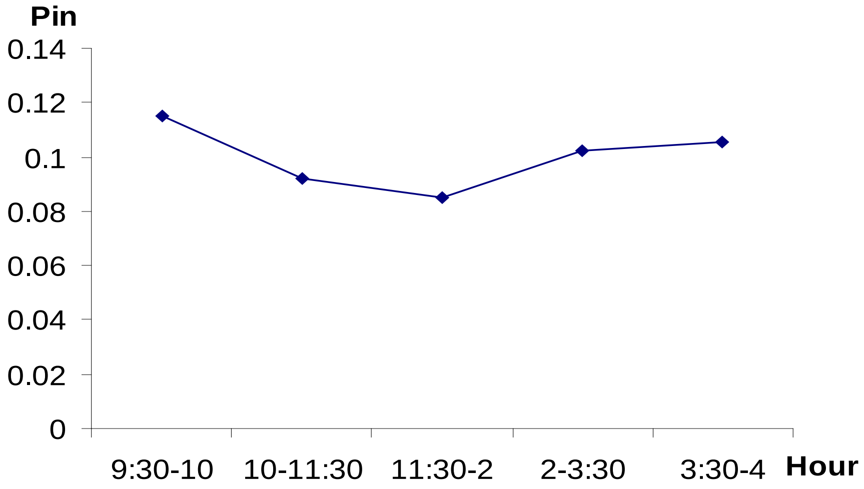
Symbol	9:30-10:00	10:00-11:30	11:30-2:00	2:00-3:30	3:30-4:00
AA	0.135	0.094	0.091	0.13	0.118
AIG	0.199	0.166	0.148	0.168	0.128
AXP	0.127	0.103	0.083	0.101	*
BA	0.119	*	0.067	*	0.098
C	0.079	0.063	0.059	*	0.11
DD	0.158	0.09	0.114	0.135	0.134
DIS	0.098	*	*	0.082	0.125
HD	0.093	0.063	*	0.063	0.062
HON	0.106	0.089	0.102	0.103	0.129
HPQ	0.093	*	0.079	0.071	0.088
JNJ	0.094	0.048	0.052	0.067	0.107
KO	0.103	*	0.089	0.093	0.076
MCD	0.139	0.135	0.131	0.144	0.1
MMM	0.082	0.085	0.086	0.091	0.107
PG	0.064	0.095	0.041	0.098	0.09
T	0.126	0.091	0.065	0.093	0.118
UTX	0.122	0.083	0.087	0.107	0.104
XOM	0.134	0.081	0.071	0.091	0.097
Average	**0.115**	**0.092**	**0.085**	**0.102**	**0.105**

PIN estimates for 18 stocks in the DJIA index. The other 12 stocks were dropped because the MLE procedure failed to converge in at least three periods. * represents missing value due to the failure of convergence.

In Figure 2, I plot the average PIN of the 18 stocks in different periods. We can see a clear U-shaped pattern there. Actually if we look at the results in Table 2, we will find that the U-shaped pattern is very consistent across stocks. The PIN in the first period, PIN_1, is bigger than the PIN in the second period, PIN_2, for 12 out of 14 stocks. PIN_2 is greater than PIN_3 for 8 out of 13 stocks, while PIN_3 is less than PIN_4 for 13 out of 14 stocks. After we perform a paired sign test, we find that $PIN_1 > PIN_2$, $PIN_1 > PIN_3$ and $PIN_3 < PIN_4$ at 1% level, which strongly supports the U-shaped pattern in PIN over the day.

Economics Bulletin, 2011, Vol. 31 No. 4 pp. 3103-3112

Figure 2 Plot of the Average Estimated Intraday PIN



Plot of the average estimated intraday PIN. The estimates are from the data for 18 stocks in DJIA from September 1, 2006 to November 30, 2006.

Our finding about the U-shaped PIN pattern over the day is consistent with Hong and Wang's (2000) idea: During the closure, market prices cease to provide information to the uninformed investors. As the informed investors continue to receive new private information, the information asymmetry between the two classes of investors increases over the night. At the open, trading resumes. The opening price partially reveals the private information of the informed investors accumulated over the closure, causing the level of information asymmetry to drop discretely. As trading continues during the day, more private information is revealed through the prices, and the level of information asymmetry may continue to decrease. At the close, two additional forces come into play. On the one hand, the informed investors cut back their hedging positions, which also unveils their speculative positions and their private information. This tends to speed the decrease in information asymmetry at the close. On the other hand, the informed investors also cut back their speculative positions due to the high risk over the closure. The reduction in speculative trade makes the stock price less revealing about existing private information. This tends to increase the information asymmetry at the close. When the second force dominates the first one, there will be a high information asymmetry around the close also.

4. Conclusion

In this article, we extend the Easley, Kiefer, O'Hara and Paperman's (1996) framework to an intraday model. We empirically estimate the intraday PINs for the 30 stocks in the Dow Jones Industrial Average Index, and find a U-shaped intraday pattern. After performing a paired sign test, we are more confident with our finding – the U-shaped PIN pattern over the time of day.
Of course, some further work can be done based on the results of this paper. For example, we can add trade size into the model and see whether the size conveys more information about the trading process. We can also try other ways of dividing the time of a day, such as every half hour, and see whether we can find a similar intraday pattern in PIN.

REFERENCES

Admati, A., and P. Pfleiderer (1988), “A Theory of Intraday Patterns: Volume and Price Variability”, *Review of Financial Studies*, 1, 3-40.

Anand, A., S. Chakravarty, and T. Martell (2005), “Empirical Evidence on the Evolution of Liquidity: Choice of Market versus Limit Orders by Informed and Uninformed Investors,” *Journal of Financial Markets* 8, 265-287.

Barclay, M. and T. Hendershott (2003), “Price Discovery and Trading After Hours”, *Review of Financial Studies*, 16, 1041-1073.

Bloomfield, R., M. O’Hara, and G. Saar (2005), “The ‘Make or Take’ Decision in an Electronic Market: Evidence on the Evolution of Liquidity,” *Journal of Financial Economics* 75, 165-200.

Chung, K., B. Van Ness, and R. Van Ness (1999), “Limit Orders and the Bid-Ask Spread,” *Journal of Financial Economics* 53, 255-87.

Easley, D., S. Hvidkjaer, and M. O’Hara (2002), “Is Information Risk A Determinant of Asset Returns?” *Journal of Finance*, 57, 2185-2221.

Easley, D., N. Kiefer, and M. O’Hara (1996), “Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow”, *Journal of Finance*, 51, 811-833.

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

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

Easley, D., N. Kiefer, M. O’Hara, and J. Paperman (1996), “Liquidity, Information, and Infrequently Traded Stocks”, *Journal of Finance,* 51, 1405-1436.

Easley, D., Engle, R. F., O'Hara, M. and Wu, L., "Time-Varying Arrival Rates of Informed and Uninformed Trades" (December 7, 2001). AFA 2002 Atlanta Meetings.

Ellul, A., C. Holden, P. Jain, and R. Jennings (2006), “Order Dynamics: Recent Evidence from the NYSE”, Indiana University Working Paper.

Glosten, L. R. and Milgrom, P. R. (1985), “Bid, Ask, and Transaction Prices in a Specialist Market With Heterogeneously Informed Traders”, *Journal of Financial Economics* 14, 71-100.

Greene, J. and S. Watts (1996), “Price Discovery on the NYSE and NASDAQ: The Case of Overnight and Daytime News Releases,” *Financial Management* 25, 19-42.

Hasbrouck, J. (1995), “One Security, Many Markets: Determining the Contributions to Price Discovery”, *Journal of Finance*, 50, 1175-1199.

Henker, T., and J. Wang (2006), “On The Imporance of Timing Specifications in Market Microstructure Research”, *Journal of Financial Markets*, 9, 162-179.

Hong, H. and Wang J. (2000), “Trading and Returns under Periodic Market Closures”, *Journal of Finance,* Vol LV, No 1, February, 297-354.

Kyle, A.S. (1985), “Continuous Auctions and Insider Trading”, *Econometrica*, 53, 1315-1335.

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

Madhavan, A., M. Richardson, and M. Roomans (1997), “Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks”, *Review of Financial Studies*, 10, 1035-1064