2011 3rd International Conference on Information and Financial Engineering
IPEDR vol.12 (2011) © (2011) IACSIT Press, Singapore

The Effect of Trading Volume on PIN's Anomaly around Information Disclosure☆

Jinfeng Ma[1+] Shancun Liu[1]

[1]School of Economics and Management, Beihang University, Beijing 100191

Abstract. Aktas et al. (2007) find lower PIN (Probability of Information-based Trading) estimates in periods with information leakage prior to the announcements of merger and acquisition and label the inconsistency as "PIN's anomaly". Using the EKOP model, this paper selects 17 stocks punished by CSRC (China Securities Regulatory Commission) for the existence of information leakage or insider trading as sample to study the reasons for PIN's anomaly based on the perspective of trading volume. The results indicate that trading volume has important effects on PIN estimation. Specifically, the phenomenon of PIN's decrease prior to information disclosures with insider trading disappears after taking trading volume into consideration. Furthermore, a new algorithm is proposed to eliminate the numerical overflow and underflow problems in PIN estimation, which sheds light on heavily traded stocks.

Keywords: PIN’s anomaly, Information disclosure, EKOP model, Probability of informed trading

1. Introduction

Science introduced by Easley et al. (1996)[1] based on a structural sequential trade model to investigate whether discrepancies in information-based trading can explain observed differences in spreads between actively and infrequently traded stocks, the EKOP model was adopted by many researchers to address a variety of issues in empirical financial studies: the PIN-return relationship[2] [3]; the information content around disclosure[4]; the role of financial analysts[5]; PIN and market efficiency[6]. Though without a direct test, the PIN measure is implicitly assumed to be an accurate measure of information-based trading in most of these studies.

To provide a direct validity test of the PIN as information based trading measure, Aktas et al. (2007)[7] investigate its behaviour around 87 cases of mergers and acquisitions (M&A) announcements that took place on Euronext Paris between 1995 and 2000. There was overwhelming evidence of illegal insider trading or information leakage prior to such events. The behaviour of the PIN around the announcement date is in direct contradiction both with intuition and with the evidence of information leakages: it decreases in the pre-event period and increases after the information release. Aktas et al. (2007) showed that the PIN’s failure to detect information-based trading was due to two major shortcomings: (1) it only reflects the number of orders, while volume is more relevant; (2) it captures the effect of public information as well as private information.

Stocks with high trading volume often suffer from the overflow and underflow problems for large buys or sells that may trigger the power function embedded in the likelihood function to generate a numerical value that exceeds the maximum or minimum range of real number values that a computer software program can handle. In this paper, a new numerical method was proposed to eliminate the numerical overflow and underflow problems in maximizing the log likelihood function of the EKOP model. We investigate the PIN

☆ This paper is supported by the National Natural Science Foundation of China (Grant No. 71071010).
+ Corresponding author. Tel.: + (01082331482); fax: +(01082338103).
E-mail address: mjf0606@qq.com

behaviour around a sample of information disclosure of 17 stocks punished by the CSRC (China Securities Regulatory Commission). First, we use trading order data in the EKOP model, and the result reproduces PIN's anomaly. Second, using the trading volume data, we obtain another measure of the probability of information-based trading, while increase prior to announcements compared to the benchmark period. These results indicate that the trading volume convey important information which is omitted in the original PIN measure.

This paper proceeds as follows. Section 2 introduces the EKOP model and our method to eliminate the numerical convergence problem. Section 3 presents our data and empirical results. We conclude in Section 4.

2. The EKOP model

2.1. Model

Based on a sequential trade model, Easley et al. (1996) developed a method to infer the probability of information-based trading using the aggregate number of buy and sell orders every day. Investors, either informed or uninformed, are assumed to trade one single risky asset and money with a risk neutral market maker. Prior to each trading day, nature determines whether an information event about the asset value take place with probability α. The orders from these traders are modelled as independent Poisson processes with arrival rates μ and ε for the informed traders and the uninformed traders, respectively. While the uninformed traders submit both buy orders and sell orders with equal probabilities on average, the informed traders submit only buy orders or sell orders according to whether the asset value revealed from the information event is high or low. The probability of arrived news to be negative is δ. Thus, there is an $\alpha\delta$ probability for a trading day to be associated with bad news, during which the informed traders submit only sell orders. Likewise, there is an $\alpha(1-\delta)$ probability that the informed traders submit only buy orders on a trading day with good news. When there is a lack of news with probability $1-\alpha$, only the uninformed traders participate in trading the stock.

In each trading day, the expected orders from the informed traders are $\alpha\mu$ while the uninformed traders contribute 2ε. Therefore, the probability of informed trading can be defined as

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \tag{1}$$

In order to obtain this proxy of the presence of informed trades, the model parameters $\Theta = \{\alpha, \delta, \varepsilon, \mu\}$ are estimated by the maximization of a likelihood function. The likelihood of observing B buys and S sells for a single trading day is

$$\begin{aligned} L(\Theta|(B,S)) = \ & \alpha(1-\delta)e^{-(\varepsilon+\mu)}\frac{(\varepsilon+\mu)^B}{B!}e^{-\varepsilon}\frac{\varepsilon^S}{S!} \\ & +\alpha\delta e^{-\varepsilon}\frac{\varepsilon^B}{B!}e^{-(\varepsilon+\mu)}\frac{(\varepsilon+\mu)^S}{S!} \\ & +(1-\alpha)e^{-\varepsilon}\frac{\varepsilon^B}{B!}e^{-\varepsilon}\frac{\varepsilon^S}{S!} \end{aligned} \tag{2}$$

Since days are independent, across the T trading days, the likelihood to maximize is

$$\mathrm{L}(\Theta|\mathrm{M}) = \prod_{t=1}^{\mathrm{T}} \mathrm{L}(\Theta|(B_t, S_t)) \tag{3}$$

Maximization of Eq. (3), with respect to the parameter vector Θ, yields maximum likelihood estimates of the parameters of interest.

2.2. The method to eliminate the numerical convergence problem

As the number of orders gets large, the likelihood function becomes harder, and even impossible in certain cases, to compute due to the factorial, the exponential and the power functions. To reduce the convergence problem of the numerical maximization of the likelihood function when the number of buys and sells is large, Easley et al. (2008) [8] suggest using the following rearranged log-likelihood function:

$$\log(L(\Theta|M)) = \sum_{t=1}^{T}[-2\varepsilon + M_t \ln(x) + (B_t + S_t)\ln(\varepsilon+\mu)] + \sum_{t=1}^{T}\ln[\alpha(1-\alpha)e^{-\mu}x^{s_t - M_t} + \alpha\delta e^{-\mu}x^{B_t - M_t} + (1-\alpha)x^{B_t + S_t - M_t}] \tag{4}$$

where M is an arbitrary constant which is usually set to $\min(B+S)+\max(B+S)/2$, and $x = \varepsilon/(\varepsilon+\mu) \in [0,1]$. Equation (4) works fairly well to alleviate the overflow and underflow problems among stocks with low to moderate trading volumes, but it is far from eliminating these problems. Stocks with high trading volume often suffer from the overflow and underflow problems even after the moderation introduced by the factorization. Given the recent trends of institutional investors breaking up their orders into smaller pieces and the increasing prevalence of high frequency traders who often submit orders of small size, more and more stocks fall into the category for which the PIN estimation simply fails.

To eliminate the numerical convergence problems, our approach to rearrange the log-likelihood function is based on the following equation:

$$\begin{aligned}\ln(a+b+c) &= \ln[e^{\ln a - \ln(b+c)} + 1] + \ln(b+c) \\ &= \ln[e^{\ln a - \ln(e^{\ln b - \ln c}+1) - \ln c} + 1] + \ln(e^{\ln b - \ln c} + 1) + \ln c\end{aligned} \tag{5}$$

The a, b, c in equation (5) correspond to the three items in the log-likelihood function (2), respectively. Given the calculation of a, b and c being substituted by the calculation of $\ln a$, $\ln b$ and $\ln c$, we convert the items similar to ε^B that induce numerical convergence problems to items similar to $B\ln\varepsilon$. Therefore very large numbers of B and S are allowed without overflow and underflow problems.

3. The data

Data used in this study is a set of high frequency trading data composed of 17 stocks punished by CSRC (China Securities Regulatory Commission) for the existence of insider trading or information leakage prior to the information disclosure. The trading data was selected from the database of Sinofin Financial Information Service Company, while the legal violations of insider trading and the dates of information disclosure were obtained from the webpage of CSRC. According to Easley et al. (1997a) [9], The PIN estimation requires a set of data of at least 60 trading days. While the announcement day being day 0, we estimate the parameters of EKOP model for each stock over three periods: [-180, -61] (the benchmark period), [-60, -1] (the pre-event period), [1, 60] (the post-event period). Table 1 provides statistics about our sample. The statistical results show that both the trading order and the trading volume undergo an increase from the benchmark period to the post-event period.

Table 1. Statistics of the Sample

		maximum	median	25th percentile	mean	st. dev.
Buy orders	benchmark	1575	478	217	516	341
	pre-event	1800	447	283	527	322
	post-event	1625	535	261	588	387
Sell orders	benchmark	1396	446	242	503	321
	pre-event	1480	450	291	538	326
	post-event	1628	532	279	574	357
Buy volume	benchmark	53189711	1631222	690533	2877435	3554369
	pre-event	42184346	1659033	720156	3056619	3717791
	post-event	37319960	1648829	647268	3153360	3838519
Sell volume	benchmark	18480139	1774988	781863	2830904	3470886
	pre-event	31924481	1785871	832690	3150498	3561456
	post-event	31280847	1869452	757803	3283929	3722444

4. The result

Using the trading order data and the trading volume data, respectively, the parameters were estimated by numerically maximizing the likelihood function given by Eq. (3). Table 2 gives the cross sectional average of the estimated parameters for each period. A classical paired test is used to determine whether the differences between means of parameters of adjacent periods are significant. *(**) represents the 10%(5%) significance level. Unsurprisingly, the parameters and the probability of informed trading (PIN) obtained from the trading order data reappear the anomaly found by Aktas et al. (2007) when investigating the M&A announcements token place on Euronext Paris. The estimated μ, the arrival rate of informed trading, dropped from 258 in the benchmark period to 245 in the pre-announcement period, which is in contradiction with the fact of information leakages or insider trading. In contrast, the estimated ε, the arrival rate of uninformed trades, increases in the last two periods. As the logical result (see Eq. (1)) of the decrease in μ during the pre-announcement period, coupled with an increase in ε, the PIN undergoes a decrease during the pre-announcement period (from 0.1220 to 0.1217). This is exactly the opposite to what has been expected.

Using the method based on equation 5 to eliminate the numerical overflow and underflow problems, We substitute B (the number of buys) and S (the number of sells) involved in the likelihood function (see Eq. (2)) by volume information (the number of shares bought or sold) to investigate the effect of trading volume on PIN. Though μ undergoes a decrease in the pre-announcement period, the anomaly of PIN during the pre-announcement period is eliminated. The value of PIN increases from 0.1525 in the benchmark period to 0.1626 in the pre-announcement period, though followed by another increase to 0.1788. The results indicate that the influence of trading volume information on PIN estimation helps to eliminate the PIN anomaly before the announcements.

It should be noted that, α, the probability of information event, increases in the pre-event period and decreases in the post-event period, which is accordant with the information leakage fact during the pre-event period.

Table 2. Statistics of Parameters

	trading-order			trading-volume		
	benchmark	pre-event	post-event	benchmark	pre-event	post-event
α	0.4259	0.4443 (4.3%)	0.4012 (-9.7%)	0.2609	0.2891 (10.8%)	0.2535 (-12.3%)
δ	0.4120	0.4861 (18.0%)	0.3705 (-23.8%)*	0.4517	0.4834 (7.0%)	0.4903 (1.4%)
ε	453	478 (5.6%)	514 (7.4%)	2429349	2669184 (9.9%)	2667539 (-0.1%)
μ	258	245 (-4.8%)	327 (33.5%)**	6153583	3516475 (-42.9%)	4469110 (27.1%)
PIN	0.1220	0.1217 (-0.3%)	0.1230 (1.0%)	0.1525	0.1626 (6.6%)	0.1788 (10.0%)

5. conclusion

Stocks with high trading volume often suffer from the overflow and underflow problems in maximizing the log likelihood function of the EKOP model. In this paper, a new numerical method was proposed to eliminate the numerical convergence problems. We investigate the PIN behaviour around a sample of information disclosure of 17 stocks punished by the CSRC (China Securities Regulatory Commission). Benefited from the new mathematical method, we managed to substitute the number of buy orders and sell orders involved in the EKOP model by the number of shares bought or sold to investigate the effect of trading volume on PIN and obtained a new measure of the probability of information-based trading, which increase prior to announcements compared to the benchmark period. The disappearance of PIN's anomaly during the pre-announcement period indicates that trading volume conveys important information and can not be omitted when estimating PIN.

It should be pointed out that the new measure of the probability of informed trading obtained from the trading volume data also undergoes an increase during the post-event period, which may be the effect of the public information disclosed in the announcements.

6. References

[1] D. Easley, N. Kiefer, M. O'Hara, and J. Paperman. Liquidity, information, and infrequently traded stocks. *The Journal of Finance*. 1996, 51(4):1405–1436.

[2] M. Kang. Probability of information-based trading and the January effect. *Journal of Banking & Finance*. 2010, 34(12):2985-2994.

[3] K. Kubota and H. Takehara. Information based trade, the PIN variable, and portfolio style differences: Evidence from Tokyo stock exchange firms. *Pacific-Basin Finance Journal*. 2009, 17(3):319-337.

[4] K. Suda and H. Takehara. Impact of Quarterly Disclosure on Information Asymmetry: Evidence from Tokyo Stock Exchange Firms. *Working paper*. 2010.

[5] D. Easley, M. O'Hara, and J. Paperman. Financial analysts and information-based trade. *Journal of Financial Markets*. 1998, 1(2):175-201.

[6] C. Vega. Stock price reaction to public and private information. *Journal of Financial Economics*. 2006, 82(1): 103-133.

[7] N. Aktas, E. de Bodt, F. Declerck, and H. V. Oppens. The PIN anomaly around M&A announcements. *Journal of Financial Markets*. 2007, 10(2):169 - 191.

[8] D. Easley, R. F. Engle, M. O'Hara, and L. Wu. Time-Varying Arrival Rates of Informed and Uninformed Trades. *Journal of Financial Econometrics*. 2008, 6(2):171 -207.

[9] D. Easley, N. Kiefer, M. O'Hara. One in the life of a very common stock. *Review of Financial Studies.* 1997, 10(3): 805-835