A Reexamination of Informed Trading and Firm Size in the Thailand Capital Market

Yafeng Qin, Bin Tang and Min Bai

Department of Economics and Finance (Albany)
Massey University, New Zealand

This version: January 3rd, 2010

Abstract

The paper investigates the stylised fact, addressed in several prior studies, that large-cap firms may have a lower probability of informed trading. Using a group of SAS codes to synthetically process 30 million intraday trades, daily trades, and firm accounting data of sample firms listed on the Stock Exchange of Thailand (SET), the paper finds an opposite pattern for the sample firms in Thailand, showing that large firms have a higher probability of informed trading. One potential reason for this finding is that large firms in Thailand tend to have high stock prices, which may repel uninformed investors; which are usually individual traders; from investing in these large firms, whereas informed investors; which are often institutional investors; may be inclined to invest in large firms. This paper provides supportive evidence for this conjecture, documenting those large companies, in general, have high stocks prices, and that high priced stocks have a higher probability of informed trading. Such a finding has important implications for publicly traded companies, investors, and market regulators, as well as policy makers.

1. Introduction

The daily stock trading process exhibits a persistent and substantial imbalanced order flow. The degree of the daily order flow imbalance is positively related to the level of informed trading (Kyle, 1985). Such trading might stem from information asymmetry, where a subgroup of investors has private value-enhancing information. The private value-relevant information may be derived from either the analysis of publicly available information, or insider information. Trading in terms of such knowledge is known as informed trading. In the context of information-based trading, traders are also divided into three groups (Bardong et al., 2007). The best school of informed investors are corporate insiders and relevant outside traders, as they should know all the value-relevant firm-specific information. The second class of investors have better ability to analyse public information, and are typically institutional traders. Finally, typical public and individual investors are uninformed investors, also known as liquidity traders, who usually bear information risk initiated by the former two classes of investor.

Information-based trading plays an important role in the security price discovery process (Kyle, 1985), and has an impact on stock prices and returns, being referred to as the information risk. Easley, Kiefer, O'Hara and Paperman (1996) develop a tool to measure the intensity of informed trading in order to explore the price discovery process and information risk. Estimated by the numerical maximisation of the likelihood function specified by a theoretical market micro-structural model using intraday transaction data, Easley et al. (1996) calculate the so-called PIN variable to depict the probability of information-based trading amongst all of the trades.

Essentially, PIN can be approximately regarded as a ratio of the daily order flow imbalance to the total daily number of order flows averaged over a certain period (Vega, 2005; Easley et al., 2008; Aktas et al., 2007). A full year horizon is often proposed to obtain a reliable value (Aslan et al., 2007; Duarte and Young, 2009). By definition, PIN can be jointly determined by the level of informed trading and the level of uninformed trading. Furthermore, the informed trading that PIN directly extracts from the order flow imbalance should be an aggregation of informed trading attributable to either the insider information, or a superior ability to process the public information.

A large body of research suggests that the degree of informed trading is associated with the characteristics of a firm. Prior studies show that firm size is negatively related to information asymmetry and, therefore, to informed trading. Easley et al. (1996) suggest that larger companies, with highly liquid stocks, probably have lower PIN than do small companies, with relatively illiquid stocks. Small firms expose public investors (uninformed investors) to a relatively high probability of informed trading than do larger firms. Several studies (Bardong et al., 2007; Aslan et al., 2007; Vega, 2005; Brown et al., 2003; Lai et al., 2008) also report similar results drawn from different samples.

In contrast, Dennis and Weston (2001) find that the activities of informed trading increase as firm size increases, since large firms are more likely to be invested in by institutional investors. Bardong et al. (2007) find that one type of informed traders, institutional investors, tend to focus their investments in large firms rather than breaking up their investments across many smaller firms.

Furthermore, Admati and Pfleiderer (1988) suggest that an increased level of liquidity trading leads to more informed trading, as informed traders act on the condition of stock liquidity and carry out a timing strategy to delay their trading until a favourable market depth emerges, trading with liquidity providers; that is, uninformed traders (Kyle, 1985). Bardong et al. (2007) further argue that more liquid stocks accompanied by possible lower trading costs should enable institutional traders to publicly trade on their private information. Therefore, an increase in informed trading is expected as liquidity increases. Easley et al. (1996) also admit that highly liquid stocks attract more informed traders than do illiquid stocks.

Easley et al. (2001) provide an explanation that, although informed trading increases as size and liquidity increase, this effect can be offset by even more uninformed traders being attracted to trade in highly liquid stocks. As a result, the proportion of informed trading to uninformed trading may be stable. Easley et al. (2001) emphasize that, other things being equal, the arrival rate of informed traders may increase as the arrival rate of uninformed traders increases. Less active stocks might even, interestingly, have a high proportion of informed trading (Hillegeist et al., 2006).

Although Easley et al. (2001) find a negative relation between the probability of informed trading and firm size and liquidity, their result is based mainly on US market data, which is the world's most developed financial market, with highly developed government regulations, disclosure policies, and sophisticated public investors. More importantly, since the Great Depression, US stock prices have remained at about USD35 per share on average, due to frequent stock splits (Weld et al., 2009). These market structure conditions may not be applicable to emerging markets. Lai, Ng and Zhang (2008) show that there is high cross-sectional fluctuation in informed trading across firm- and country-level characteristics. This implies that some empirical results derived from the US market may not necessarily be valid for other financial markets, especially the emerging markets.

The Thai stock market is a typical emerging market, dominated by individual investors, which account for 70% of the daily market trading volume (Pavabutr and Sirodom, 2007). These public investors' investments, however, mainly concentrate in stocks with low prices, in particular below THB5. Of these low-priced stock trades, 90% stem from public investors, whereas for stocks above THB 200, institutional investors account for 76% of all transactions (Pavabutr and Sirodom, 2007).

Interestingly, small firms in Thailand tend to have low stock prices, whereas large firms usually have high stock prices (Pavabutr and Sirodom, 2008). This means that stock price is an increasing function of stock market capitalisation. This market structure is very different from the US market. Large firms in Thailand, with higher stock prices, may incur clientele problems. Thai individual investors, who are also usually uninformed investors, may shy away from stocks with high prices (Copeland, 1979). As uninformed investors are more wealth-constrained, they will prefer to buy lower priced stocks which are associated with smaller ticker size and, thus, lower transaction costs in absolute values (Baker and Gallagher, 1980). Lower stock prices also provide marketable and affordable round lot trading (Lakonishok and Lev, 1987). The monotonic positive relation between firm capitalisation and stock prices will prevent Thai public investors from investing in large firms. On the other hand, a lower stock price means lower minimum price increments, and higher proportional bid-ask spreads. Institutional investors, more likely to be informed traders, must pay higher transaction costs when they fulfil block trades of such low price stocks (Falkenstein, 1996). Furthermore, to avoid the price impact, they also need to mask their trading purposes more carefully by breaking up their block trades into smaller sized trades, leading to an increase in commission and monitoring costs (Bernhardt and Hughson, 1997; Chordia and Subrahmanyam, 2004; Barclay and Warner, 1993; Chakravarty, 2001; Hasbrouck, 1991, 1991a, 2007). Under the Thai stock market characteristic of the monotonic positive relation between firm size and firm stock price, institutional investors may have an adverse stock-picking strategy for information-based trading.

Furthermore, the Thai stock market is a relatively open market, that allows foreigners to invest in Thai securities settled in Thai Baht, with only a modest restriction. It is worth noting the difference in the profiles of local and foreign investors, in that local holdings are typically individuals, while foreigners are typically sophisticated institutional investors. Try to image that a Thai stock with a 50% maximum foreign ownership restriction begins to trade on the Alien board. This may imply that about 50% of the firm shares are likely to be held by foreign investors, who are mainly institutional investors. It is, therefore, plausible to assume that foreign institutional investors must exert a broad influence on the Thai stock market, which has a special dual market regime. Similarly, some researches (Choe et al., 1999) document that domestic institutions generally behave like foreign investors. Ferreira and Matos (2007) find that all institutions have a strong preference for large and liquid stocks with good governance practices. Covrig et al. (2006) suggest that domestic institutions, which are also informed investors, have a similar, but slightly wider, range of investment preferences relative to international institutional investors.

Merton (1987) indicates that, if having only incomplete information on security, "…an investor uses security in constructing his optimal portfolio only if the investor knows about the security". Banz (1981) and Reinganum (1981) also find that large securities, which usually have more publicly available information, are preferred by institutional investors.

Aslan et al. (2007) further argue that, unlike the prior literature, "…firm size is not simply an information effect, but rather may arise for a variety of factors related to small firms such as financial constraints, bankruptcy concerns, or the like". For investors with incomplete information, firm capitalisation means not only information availability, but also visibility. Falkenstein (1996) provides supportive evidence that informed investors bias their investments to certain stocks with large capitalisation, high liquidity, and more visibility. This could be an explanation of why international institutional investors do not invest in relatively small firms with few stockholders.

For the Thai stock market, small Thai firms are usually opaque and high ownership concentration. International investors have great difficulty in gathering their value-enhancing information for these poorly-known Thai firms. Bailey, Mao and Sirodom (2006) suggest that foreign investors on the Thai stock market are usually sophisticated institutional investors with better information-processing ability and, therefore, the trading they initiate is more likely to be information-based. Furthermore, foreign institutional investors tend to devote their valuable information processing skills to assessing large, less price impacted firms. Bailey and Jagtiani (1994) conclude that foreign investors prefer investing in large well-known Thai firms. Based on Merton's Investor Recognition theory, the international investors may find it easier to obtain and process information on those large, and hence information-rich, Thai firms.

Consistent with Easley et al. (2001), Brown, Hillegeist and Lo (2006) also suggest that large firms with better visibility attract uninformed investors. As discussed above, in terms of the Thai stock market, however, these visible large cap stocks may have too higher stock prices to encourage individual Thai investors to them. As a result, when choosing investment targets, informed investors may prefer large, highly visible firms, whereas uninformed traders may desire small cap companies. Hence, I expect that the mismatched trader composition may cause some deviation from the negative relation between firm size and the probability of informed trading shown in prior studies.

Falkenstein (1996) demonstrates that informed traders, like institutional investors, may comprehensively assess their investment on the basis of liquidity and information asymmetry. For international investors, this may involve a trade off between the marginal benefits and the marginal costs of gathering and processing value-enhancing information. Fortunately, large and visible Thai firms are generally inexpensive in terms of the market to book ratio, which usually ranges from one to three, being significantly lower than in many other developed and developing markets. Other things being equal, informed traders would rationally choose larger and well-known firms in the hope of improving their profits per unit of risk.

In short, since large firms in Thailand are often highly profitable and growth option companies with a low price impact and low stock price volatility, yet have relatively high stock prices and low turnover, these large companies are ideal investment targets for institutional investors, particularly international investors. Given the above discussions, I am especially interested in exploring how informed trading displays itself in the Thai stock market.

For other emerging markets, Ciner and Karagozoglu (2008) also find direct evidence that speculation based on informed trading is more prominent for larger firms, which contradicts the findings from the US market.

Chan, Menkveld and Yang (2008) also suggest that, in China, the highly liquid A-share market is known for its abundant information-based trading, especially in terms of insider trading and block trading, relative to the less liquid B-share market.

All of these controversies motivate me to explore the following question:

Does the negative relation between firm size and informed trading suggested by Easley et al. (1996) always exist in all markets?

Testable hypotheses:

Based on Thai firm samples, large firms have a higher probability of informed trading than do small firms.

Based on a set of Thai firm samples, firms with high prices have a lower level of uninformed trading.

The remainder of the thesis is organised as follows. Section 2 introduces the informed trading measure (PIN), what relation may exist between informed trading and firm size, some disadvantages of PIN, and why AdjPIN is preferred by some researchers. Section 3 presents the methodology and how to determine the essence behind these informed trading measures. Section 4 details the data, provides a brief review of the Thai stock market, and introduces the data processing procedures, the PIN and AdjPIN estimations, and the control variables selection. Section 5 reports the empirical results and the analysis of these results. Section 6 provides a discussion on some possible explanations and summarises the paper.

2. Related Literature

2.1 Informed trading and PIN

Informed trading cannot easily be discriminated from routine market activities. One possible way to do so is to observe abnormal imbalance order flow and indirectly infer the probability of such strategic trading hidden in day to day trading (Chordia and Subrahmanyam, 2004). This intuition has been adopted by Easley et al.'s (1996) information asymmetry model. Using a nonlinear likelihood function, Easley et al. (1996) developed a percentage measure to demonstrate the ratio of informed investors to uninformed investors (liquidity investors) in a certain stock.

By using this measure, they find that the market makers for NYSE traded stocks generally give more weight to informed trading; that is, imbalance order flow; rather than the trading volume when deciding the bid-ask spread. They also admit that high trading volume stocks have higher numbers of informed traders than do low trading volume stocks. They argue, however, that this effect can be offset by even more uninformed traders, who may be attracted to trade high trading volume stocks. Consequently, they note that less active stocks might, interestingly, have a high proportion of informed trading, largely in the form of insider trading.

Although information asymmetry, by definition, is a determinant of informed trading, we need to clarify some issues. Bardong et al. (2007) analyse the sources of information asymmetry, either due to value-enhancing information derived from the analysis of publicly available information, or related to insider information. Although PIN directly captures the intensity of informed trading, its outcome can reflect various aspects of informed trading, which might be based on either advance knowledge of the proprietary industry, or macro forecasts, insider information, and a superior ability to process public information.

As to firm size, Bardong et al. (2007) state that institutional investors tend to focus their investments in large firms, instead of spreading their investments across many smaller firms, as these investors are more experienced than public investors. The most likely reason is that more liquid stocks, accompanied by possible lower trading costs, should enable institutional traders to publicly trade on their private information. Therefore, informed trading is expected to increase as liquidity increases.

Dennis and Weston (2001) also discuss a similar relation. Their paper notes that since institutions are better informed than individuals, intuitional investors implement more speculative trading on firms with high levels of institutional ownership.

Easley, O'Hara and Saar (2001) repeatedly emphasize the idea in a series of their papers that information trading is a dynamic game. Other things being equal, the arrival rate of informed traders may increase as the arrival rate of uninformed traders increases, leading to a relatively stable information-based trading ratio, even if the firm's trading condition may have changed, such as when a stock split take places. This is more likely to be an optimal problem in game theory.

Easley et al. (2002) depicts that PIN only accounts for the degree of informed trading, rather than the influence of these trades on excess returns, the relation between PIN and other firm proxy variables should be expected to vary in terms of the sample. This may make the relation an empirical issue.

PIN can be directly calculated from trading data, and can creatively facilitate the pattern of order flow to trail the informed trading, then identifying the information risk. PIN is, however, difficult to connect to the firm-specific variables. Aslan, Easley, Hvidkjaer and O'Hara (2007) recently explicitly linked PIN to firm accounting and characteristic measures. They attempt to determine a set of proxy firm specific variables to represent unobservable PIN, thus overcoming the obstacle of computing PIN and simplifying the usage of PIN.

Aslan et al. (2007) believe that information risk due to private information can be statistically represented by a variety of firm accounting natures, such as analyst coverage and abnormal accruals. This set of proxy variables can be seen as the information environment of a firm. Aslan et al. (2007) regress a time-series of PIN on a wide variety of specific firm and industry variables, and find some interesting relations between informed trading and industry structure and firm characteristics such as age, size, growth, profitability, insider holding, institutional trading, and accruals. The paper again confirms that firm size has a negative relation with PIN, indicating that larger firms have less informed trading.

When incorporating PIN into the Fama and French pricing model, they find the most significant explanatory variable is still firm size. They further argue that, unlike the extant literature, "firm size is not simply an information effect, but rather may arise for a variety of factors related to small firms such as financial constraints, bankruptcy concerns, or the like" (Aslan et al., 2007). They also confirm that informed traders generally seek out high quality investment targets in terms of high potential returns, that institutional investors are more likely to be informed, and that firms with high institutional holdings are subject to high levels of informed trading.

Since PIN was first published, a large number of papers have used PIN to study a broad range of topics in corporate finance, investments, and market microstructure. Employing the PIN measure, Brown, Hillegeist and Lo (2004) examine whether firms' conference calls can effectively lower information asymmetry and, hence, reduce the cost of capital. They again prove that informed trading is highly correlated with firm characteristics and information production policy, since conference calls can be viewed as improving firm visibility, and lowering information asymmetry. They also show, however, that although improving firm visibility attracts uninformed investors, informed investors simultaneously increase due to firm quality improvements. As a result, the ratio of informed trading to liquidity trading may be stable. The authors conclude that, in practice, informed traders such as institutional investors might be likely to have more constraints than uninformed investors, typically individual investors, would have. Consequently, the net influence of conference calls may present a decrease in PIN, because of a disproportional increase in uninformed investors.

Brown, Hillegeist and Lo (2006) study how PIN changes when a firm meets, or beats, its earning expectations, and a price premium is observed due to a reduction in the cost of capital. Meeting expectations affects investors' trading decisions and information search activities. This finding is consistent with the Investor Recognition Theory proposed by Merton (1987), which suggests that good earnings performance improves a firm's visibility. They also provide support for firm visibility attracting uninformed investors.

Vega (2005) studies the Post-earnings Announcement Drift (PEAD) via PIN. She provides deeper insight on what informed trading refers to, and the drive behind abnormal order flow, the underlying data from which PIN is calculated. Vega (2005) also states that PIN originated from trade data is not simply a measure of insider trading, but also captures informed trading initiated by investors who have superior ability to analyse private components of public information, even given public information, and are very skilful in predicting future macroeconomic activities, such as foreign exchange and bond investments.

2.2 Information context of trading and investor preferences

A stream of academic papers study the firm size effect. As well as Fama and French (1993), Banz (1981) and Reinganum (1981) also document that, based on stocks from US markets, small firms tend to have higher risk-adjusted returns than do large-cap firms. When Barry and Brown (1985) examine an equilibrium pricing model along with a loosened assumption that relatively little information is accessible for some stocks in comparison to that available for other stocks they find, however, that the firm size effect can be explained by firm listing age. They find that those small firms with a long period of listing do not produce abnormal returns. This might imply that it is the period of listing, which is a proxy of information and visibility, rather than firm size that imposes an influence on the stock return.

Merton (1987), on the other hand, develops a capital market equilibrium model, which relaxes the assumption of investors having an equal information set. He states that information gathering and diffusion should play a key role in all areas of finance, as '…an investor uses security in constructing his optimal portfolio only if the investor knows about the security' (Merton, 1987). This could be an explanation as to why some institutional investors do not invest in relatively small firms with few stockholders.

Falkenstein (1996) provides striking complementary evidence about the relation between institutional investors' preferences and firm size, and information asymmetry, which can help with the understanding of firm size in the capital market. He finds that mutual funds bias their investments to certain stocks with large capitalisation, high liquidity, and greater visibility. Large firms may also represent lower transaction costs, more liquid investments, and by allowing investors to take a large position for such stocks without a significant price impact.

Falkenstein (1996) also puts forward an interesting observation that stocks with a lower price may have a higher bid-ask spread ratio over the average price. The concern regarding transaction costs gives institutional investors an aversion to low priced stocks, and a preference towards high priced stocks.

To examine differences in investment preference, Gompers and Metrick (2001) simply divided the investors into two groups; institutional investors, and individual investors. They find that institutional investors demand stocks that are larger, more liquid, and have had relatively low returns. In addition, institutional investors might care more about the transaction costs, due to the large relative spread for illiquid or low-priced stocks, when they establish and turn around a trading position. They also pointed out that well-established large firms may have more outstanding performance, making them more appealing.

Dahlquist and Robertsson (2003) find similar preferences for Swedish firms. They further report that foreign investors tend to underweight firms with ownership concentration. The finding may have an important implication for the Thai market, since Thai publicly listed firms are mainly family controlled, especially in the case of small firms (Claessens et al., 2000). Dahlquist and Robertsson (2003) also provide evidence that foreign investors are typically mutual funds, or other institutional investors, and that domestic institutional investors tend to have the same preferences for large firms as do foreign institutional investors. Both types of investors deviate from market portfolios, indicating a profile of more informed and sophisticated trading by these investors. Dahlquist and Robertsson (2003) also demonstrate that individual investors have a preference for smaller firms with high ownership concentration, and an aversion to large sized firms with a low turnover.

Covrig, Lau and Ng (2006) examine the similarities and differences between domestic and foreign fund managers from 11 developed markets. They identify that both types of investor are equally experienced, however, they might not be equally informed. If so, the foreign fund managers would be likely to invest primarily in stocks with high visibility and worldwide recognition, or stocks that can easily lower information-gathering costs, while domestic managers put more emphasis on the fundamental characteristics of the target stocks.

In contrast to institutional investors, who prefer stocks with high prices due to the lower transaction costs, individual investors prefer buying round-lots of low-priced shares. Copeland (1979) argues that, though low priced stocks with a high relative spread may hurt institutional investors, it is somewhat cheaper to diversify, because smaller values can be traded at round-lot transactions costs.

Baker and Gallagher (1980) survey stock splits. They reveal that a stock split is usually used to bring the stock price to an optimal trading range. A lower stock price will attract investors and then enhance the ownership base. This kind of clientele that is attracted by the lower price after a stock split is usually thought as the uninformed, or small investors.

Lakonishok and Lev (1987) use the monthly turnover ratio to examine a stock's marketability. They find that the stock spilt does not permanently affect the volume of trade, and infer that lower stock prices might influence other aspects of marketability, such as the composition of stockholders.

2.3 The Thai stock market and other emerging markets

Bailey and Jagtiani (1994), in an extension of Bailey's studies of emerging financial markets, especially the Thai market, examine the stock price premium between the Main board and the Alien board in the Thai stock market. After analysing the cross-sectional variation in the price difference over firm size, they find that the Alien board price premium is higher for large firms and, thereby, conclude that foreign investors prefer to invest in large well-known Thai firms. Based on the Merton Investor Recognition Theory, international investors may find it easier to obtain and process information on information-rich Thai firms. Furthermore, foreign investors will give more weight to liquid stocks, especially those international institutional investors who usually duplicate a capitalisation-weighted country portfolio.

In another recent paper, Bailey, Mao and Sirodom (2006) shed some light on this area of research. They point out that, on the Thai stock market, foreign investors are usually sophisticated institutional investors with better information-processing ability. Therefore, the trading they initiate is more likely to be information-based. Furthermore, foreign institutional investors tend to devote their valuable information processing skills to assessing large, liquid firms, in line with my hypothesis. They also discuss a multiple markets trading behaviour where investors in the two markets, the Main board and the Alien board, may observe the stock price trend in one market and simultaneously make the trend spread (contagion) over and arise in the same direction in the other market. To a degree, this can enhance the price movement pattern in a certain stock. These evidences support my finding that foreign institutional investors may exert an effect on the Main board through contagion over both markets.

Bailey, Mao and Sirodom (2008) find that, although there is a dual board regime in Thailand, cross-market traders are very large and play a role in price discovery. These cross-market traders are typically informed traders and their trading may make profits, even though these trades are costly themselves. This evidence enhances my finding that foreign investors may also have a cross-market trading channel to capitalise their value-enhancing information, because their trading accounts for 15% of the Main board trading, and is almost certainly informed trading.

Based on the PIN measure, Lai, Ng and Zhang (2008) employ a Taqtic dataset to test the pattern and extent of informed trading in 42 countries, across 27,042 firms, and over a period of 12 years. They show that there is much cross-sectional fluctuation in informed trading across firm- and country-level characteristics.

Using market microstructure models with PIN, Chan, Menkveld and Yang (2008) explore the information asymmetry effect on equity prices in China's local A- and foreign B-share markets. This paper gives a striking, but completely opposite example regarding the relation, which suggests that stocks with high illiquidity are accompanied by high information trading. Alternatively, in China, the highly liquid A-share market is known for its abundant information-based trading, especially insider trading and block trading, relative to the less liquid B-share market.

Ciner and Karagozoglu (2008) report some empirical evidence on the relation between informed trading and foreign trading activity on the Istanbul stock market. One of the intriguing findings of this paper is that the larger the firm size, the higher the probability of informed trading that the firm is subject to. They conjecture that this might result from a lower percentage of public ownership in these large firms. Ciner and Karagozoglu (2008) also document that local investors in Turkey similarly perceive foreign investors as being experienced investors and, to a large extent, they will follow the international investors' trading strategies, making the market more volatile.

2.4 Drawbacks of PIN and its extended measure, AdjPIN

Since the PIN measure was developed in 1996, a branch of studies has raised some doubts as to its effectiveness. Aktas et al. (2007) examine the measure, using a set of 87 French firms involving merger and acquisition announcements from 1995 to 2000. They find that PIN decreases prior to the public announcement day and increases after the announcement, a phenomenon which is inconsistent with previous studies about M&A transactions, which are known to foster a high probability of insider trading, a strict form of information-based trading.

Aktas et al. (2007) argue that PIN is merely seen as the ratio between the expected absolute order imbalance (absolute difference between purchases and sales, referred to as OIB) and the expected total number of trades, since the daily PIN is able to be simply and empirically approximated by the daily relative order imbalance, the ratio between the daily imbalance and the daily order volume.

They debate that PIN fails to effectively forecast the pattern in their sample, and attribute this failure to some major drawbacks. PIN may miss improperly the volume (the number of shares recorded in each identified transaction) and the value (the executed price associated with these transactions) dimensions of the trading process, which should contain information about informed trading. Moreover, PIN may capture not only private information, but also reflect heterogeneous beliefs among investors, global trends, and the like. PIN also neglects the fact that imbalanced order flows are time-varying, and this sort of fluctuation may also disclose the presence of private information (Aktas et al., 2007).

In a recent working paper, Benos and Jochec (2007) raised four doubts about the reliability of the PIN. Apart from ignoring the number of shares in each transaction, PIN is also thought to obscure the difference between the concepts of information asymmetry and informed trading. Contrary to the assumption of the underlying PIN, any information event should span a number of days, and informed traders may act on the information over a relatively long horizon.

In addition, Venter and de Jongh (2004) find that PIN also fails to describe the real trading process. To calculate PIN, an independent Poisson distribution outlining the daily behaviours of buy and sell order flow needs to be presumed, with the expected daily arrival rates of uninformed buy and sell orders assumed to be a constant. These daily numbers are orthogonal with each other. Venter and de Jongh (2004) find, however, that these premises are unrealistic, since any information event, including both public and private information, influences the daily arrival rates of uninformed buy and sell orders, and shows a positive correlation.

Brown et al. (2006) accept Venter and de Jongh (2004) finding, and also detail that the fundamental assumption that daily arrival rates of uninformed buy and sell orders drawn from independent Poisson distributions lead the daily rates of uninformed buys and sells to be uncorrelated, is not realistic. This is because, in the real world, public events can simultaneously impact the trading intensity of buy and sell orders, such that "…then the daily arrival rates of uninformed buy and sell orders will be positively correlated" (Brown et al., 2006). Instead of a standard PIN, they use a distribution assumption provided by Venter and de Jongh (2004).

Duarte and Young (2009) find that PIN is not only a proxy for information asymmetry, but also contains a component of illiquidity that is not related to information asymmetry. Based on these findings, Duarte and Young (2009) extend the PIN measure developed by Easley et al. (1996) to accommodate the positive relation. Their extended model allows for simultaneous positive shocks on both buy and sell order flows, which increases the high volatilities of both buys and sells. For the purpose of accounting for liquidity in the spirit of the PIN measure, these authors modify PIN and separate it into two components; AdjPIN, which is associated with asymmetric information, and PSOS, which is a proxy of illiquidity.

Chordia, Huh and Subrahmanyam (2006) examine the relation between trading activity, measured by turnover, and firm visibility, represented by firm size, age, price, and the book to market ratio. The paper shows that the order flow imbalance can result simply from illiquidity, rather than from informed trading, which is consistent with the intuition behind AdjPIN.

In a recent paper, Easley, Engle, O'Hara and Wu (2008) finally admit that PIN can be viewed as a simple measure of liquidity. They examine the property of time varying in the arrival rates of both informed and uninformed trading orders, and extend the original model to incorporate the time fluctuation of order flows into a GARCH model. After analysing a set of data covering 16 randomly selected stocks that are actively traded on the NYSE, they conclude that uninformed trading is more volatile than informed trading, and that both types of trade follow some correlated, autoregressive dynamics.

3. Methodology

3.1 Measures of intensity of informed trading

Easley et al. have spent decades since 1990s developing and extending the PIN measure in order to detect informed trading from the perspective of market microstructure. Symbol α denotes the rate of an information event occurring. When an information event occurs, it might be bad news concerning information regarding the decreased value of an underlying asset with a probability of δ, or good news with the probability of 1- δ. The arrival rate of informed traders is denoted as μ.Trades are assumed to arrive independently at the market in terms of the Poisson process. ϵ_b denotes the arrival rate of uninformed buyers. ϵ_s is the arrival rate of uninformed sellers. All of the above can be summarised as one of three Poisson processes, which is estimated using MLE to obtain a parameter vector θ=($\alpha, \delta, \epsilon_b, \epsilon_s, \mu$) .

$$L(B, S|\theta)=(1-\alpha)e^{-\epsilon_b}\frac{\epsilon_b^B}{B!}e^{-\epsilon_s}\frac{\epsilon_b^S}{S!}+\alpha\delta e^{-\epsilon_b}\frac{\epsilon_b^B}{B!}e^{-(\mu+\epsilon_s)}\frac{(\mu+\epsilon_s)^S}{S!}+\alpha(1-\delta)e^{-(\mu+\epsilon_b)}\frac{(\mu+\epsilon_b)^B}{B!}e^{-\epsilon_s}\frac{\epsilon_s^S}{S!}$$

(1)

In Equation 1, B and S stand for the total daily number of buyer- and seller-initiated trades, respectively. The PIN represents the expected ratio of trades arising from informed trades, to the total order flow. The fraction can be simply written as:

$$PIN=\frac{\alpha\mu}{\alpha\mu+\epsilon_b+\epsilon_s}$$

(2)

Duarte and Young (2009) extend Easley et al.'s (1996) PIN measure to allow for the large buy and sell variance, and test the positive correlation between buy and sell orders.

Duarte and Young (2009) relaxed the strict assumption in Easley et al.'s (1996) model by allowing the arrival rate of informed buyers μ_b to be different from the arrival rate of informed sellers, μ_s . This relief can better fit the fact that buy trades are more volatile than sell trades. Moreover, their model takes into account the symmetric order flow shock. They use θ to represent the probability of the situation, conditional on there being no arrival of private information, and θ′ to denote the probability conditional on the arrival of private information. The additional arrival rate of buys and sells are Δ_b and Δ_s, respectively. They also design an adjusted probability of informed trading, being:

$$\textbf{AdjPIN} = \frac{\alpha \times [(1-\delta) \times \mu_b + \delta \times \mu_s]}{\alpha \times [(1-\delta) \times \mu_b + \delta \times \mu_s] + (\Delta_b + \Delta_s) \times [[\alpha \times \theta' + (1-\alpha) \times \theta] + \epsilon_b + \epsilon_s} \quad \textbf{(3)}$$

Duarte and Young (2009) concurrently develop a measure to account for the probability of a symmetric order flow shock denoted as PSOS. The probability is represented as:

$$\textbf{PSOS} = \frac{(\Delta_b + \Delta_s) \times [[\alpha \times \theta' + (1-\alpha) \times \theta]]}{\alpha \times [(1-\delta) \times \mu_b + \delta \times \mu_s] + (\Delta_b + \Delta_s) \times [[\alpha \times \theta' + (1-\alpha) \times \theta] + \epsilon_b +} \quad \textbf{(4)}$$

After conducting a series of comparisons and Monte Carlo simulations, Duarte and Young (2009) select a preferred extended model by restricting θ= θ′.

3.2 How to understand the economic intuition behind PIN and Adjpin

It is critical to my research to understand what the PIN, or AdjPIN, specifically refers to, since my finding is the opposite of the theoretical fact that PIN, or AdjPIN, is negatively related to firm size.

From the Equation 2, we can unambiguously find that the value of PIN can be jointly determined by the frequency of private information events (α) , the frequency of informed trading (μ), and the arrival rates of uninformed trades. In short, the intensity measure is not only literally a measure of informed trading, but also a metric of uninformed trading.

If we take a closer look at these measures, though they use a non-linear statistical technology, it can be seen that they actually have straightforward meanings. Easley et al. (2008) give an explicit explanation on what these parameters means. The parameter α represents the percentage of days with abnormal order flows. The parameter δ denotes the ratio of the number of days with net abnormal selling trades relative to the total number of days that have net abnormal selling trades relative to the total number of days that have abnormal trades. The parameters ϵ_b **and** ϵ_s represent the daily average number of buy and sell trades, respectively. The parameters μ_b **and** μ_s are interpreted as the number of independent abnormal sell and buy trades, respectively.

Similarly, the parameter θ reflects the number of days with an average change in the abnormal order flow divided by the number of days without abnormal trades. The parameters Δ_b **and** Δ_s are the daily average additional number of buy and sells, respectively.

The expected total number of trades per day, E [B+S], is equal to $\alpha\mu + \epsilon_b + \epsilon_s$. The expected daily abnormal order flow E [B-S] is equal to $\alpha\mu \times (1 - 2\delta$). The expected daily absolute abnormal order flow E [|B-S|] is expressed as $\alpha\mu$ **.** Finally, the PIN can be interpreted as:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \epsilon_b + \epsilon_s} = \frac{E[B-S]}{E[B+S]}$$

(5)

From the equation, we can again justify that the PIN is simply a description of the relative strength, or pressure, of the abnormal order flows to total order flows. In effect, the ratio is simply an ex post consequence and does not uncover the causality of the evolution of abnormal order flows. Uninformed investors are often public individual investors. They usually have unsophisticated trading strategies, which may show a passive nature. Some empirical findings from the US market suggest that informed investors, who are usually sophisticated investors, may observe the trading behaviour of uninformed and unsophisticated investors to decide on their own trading strategies in the hope of maximising profit (Kyle, 1985). This regularity may imply that the PIN ratio is mainly determined by the trading willingness on a firm stock of informed investors in terms of overall conditions of the trading environment and structural characteristics of a firm (Bardong et al., 2007), but does not remain a fixed relation with certain characteristics of a firm.

Cross-sectional regressions

To empirically explore the relation between the measure of informed trading, PIN and AdjPIN, and firm size, involves a cross-sectional test. The other variables associated with information-based trading may confound the effect of firm size on variations in PIN and AdjPIN. It is necessary to conduct a robustness test by introducing control variables. The regression equation is:

$$\boldsymbol{Pin_i}\text{(or}\boldsymbol{Adjpin_i}\text{)} = \boldsymbol{\alpha_1 + \alpha_2 size_i + \alpha_3 Amihud_i + \alpha_4 Roa_i + \alpha_5 Stdev_i + \alpha_6 Tobinq_i + \varepsilon_i} \quad \textbf{(6)}$$

In Equation 6, size represents the natural log of the market value of equity in firm i. Amihud refers to the Amihud measure, an illiquidity measure that defines an average ratio of the daily absolute price percentage change over the trading volume in dollar value on that day in firm i. Roa represents the return on assets in firm i in the last year. Stdev is the annualised standard deviation of daily returns for firm i in the last year. Tobinq is the Tobin'Q, which is obtained by the market value of equity plus the book value of liabilities in firm i divided by the book value of the total asset of firm i.

4. Data

4.1 Data collection

I choose intraday transactions and the limit order book data for 2006 on the Stock Exchange of Thailand. The data are obtained from the Securities Industry Research Centre of Asia-Pacific (SIRCA) for the purpose of calculating yearly-firm AdjPIN and PIN. As the intraday data (up to 30 million) is too large to manually calculate, I successfully programme a group of SAS codes to integrate and process these data sets.

In my research, financial institutions are excluded from the sample and only stocks that have at least 120 days of trade, or quote, data available are selected. A sample covering 58 firms remains in my research.

The Market Value (MV), daily trading volume in quantity of shares traded (VO),, daily trading volume in value (VA), net income (wc01751), total assets (wc02999), liabilities (wc03351), the market to book ratio (MTBV), the daily price (P#S), and the number of shares outstanding (NOSH) are retrieved from Data Stream.

4.2 Overview of Thai stock market

The Security Exchange of Thailand (SET) is a pure order-driven, automated stock exchange, which has no designated market makers. Investors can only put limit orders, which are sequentially arranged in an electronic open limit order book.

Trades from individual investors, which are also local investors, account for 70% of the daily market trading volume. In particular, over 90% of trading below THB 5 is most likely to be initiated by individual investors. Stock price tends to be an increasing function of stock market capitalisation. This means that large firms usually have high stock prices (Pavabutr, 2008).

Furthermore, the Thai market is artificially separated into two boards to restrict foreign ownership, The Main board is open to locals, and the Alien board is open to foreigners.

A particular stock typically trades at higher prices on the Alien Board than it does on the Main board.

Foreigners have a strict ownership limit for Thai firms set at a maximum of 50% of the shares. If the ownership of a stock has reached that limit, then foreign buyers can only obtain stock in the Alien board from a foreign seller, even though the same shares listed on both boards have identical claims on the firm.

It is worth noting the different profiles of local and foreign investors. Local shareholders are typically individuals, while foreigners are typically sophisticated institutional investors.

4.3 Classification trades and Data cleaning

4.3.1 Classification trades

To implement the Maximum Likelihood Estimation (MLE) required for computing the PIN and AdjPIN, the daily number of buy- and sell-initiated transactions need to be estimated. The key element in counting the transactions is to identify the direction of each trade. As the SET (Stock Exchange of Thailand) has no designated market maker, I define the quote spread as the difference in the best selling and buying limit orders.

Following a technique developed by Lee and Ready (1991), I classify a transaction as buyer-initiated (seller-initiated) if the transaction is above (below) the mid-point of the quoted spread. If the trade occurs at the mid-point of the spread, I classify this sort of transaction as buyer-(seller-)initiated in terms of a previously quoted mid-point spread. The transaction can be classified as a buy if the executed price is higher than the previous quoted mid-point spread, and as a sell if it is lower than the previous quoted mid-point spread. If the transaction coincidently occurs again at the same price as the quoted mid-point spread, the transaction price will be compared to the further lagged quoted mid-point spread until the value has been lagged three times.

4.3.2 Data cleaning

Trades

SIRCA provides all of the intra-day stock trades' data required for this examination, including ticker, date, time, type indictor of quote or trade, trade price, best bid price, best ask price, and qualifiers. Qualifiers can roughly identify the trades at open and close time, and can also provide a mark on doubtable and irregular trades. Using the qualifiers, I find it useful to remove the first and last 30 minutes of each trading session from the sample. The first trade each day is excluded from our sample using qualifiers. Trades with non-positive prices or sizes, or that are greater than 150%, or less than 50%, of the price of the previous trade, are also eliminated from the sample. Using a filter coded with SAS software, I consolidate all trades occurring within each five second period as a single trade, and record the executed time as the time of the first of these trades.

Quotes

The prevailing quote to be compared with the transactions must be five seconds prior to the time when a trade is executed. Otherwise, I use the previous quote for the classification scheme. The bid price must be lower than the ask price, and both ask and bid prices must be positive. The bid-ask spreads calculated from the quote also need to be less than half, or 25%, of their mid-point quote prices, otherwise I remove them.

Trading days

I exclude firms for which the average daily number of buys or sells is less than 60, or where there are either no buys, or no sells. Zero volume trading days are also eliminated. I select those firms with more than 120 trading days of both quotes and trades.

Matching with accounting data

I also excluded stocks for which accounting data are not available. After cleaning the transaction data, I merge the qualified stocks with the DataStream database to obtain the firm accounting information, using the RIC (ticker) codes provided by SIRCA. As a result, I get the most paring data sets in 2006.

4.4 PIN and AdjPIN estimation

Using the daily summation of buys and sells, I estimate the yearly-PIN and AdjPIN measures, with a maximum likelihood function for each firm, in 2006. I choose a Newton-Raphson method to maximise a log form of the joint likelihood function as:

$$L(\theta|M) = \prod_{i=1}^{t} L(\theta|B_i, S_i) \tag{7}$$

Where $\theta=(\alpha, \delta, \epsilon_b, \epsilon_s, \mu)$ for Pin, and $\theta=(\alpha, \delta, \epsilon_b, \epsilon_s, \mu_b, \mu_s, \theta, \Delta_b, \Delta_s)$ for for AdjPIN. Given that the probabilities of each trading day are independent, the joint likelihood can be calculated as a product of the daily likelihoods.

To avoid corner and local optimal solutions, I run each joint maximisation algorithm, starting from more possible and different initial values, and obtain a satisfying coverage of my sample. Fortunately, the overflow problem due to factorisation of the large value of buys and sells is not a question for the Thai stock market, since Thai stocks are all traded at a modest level.

I exclude solutions with all parameters of PIN and AdjPIN equal to zero, or one. This excludes only 6 of the 72 firms.

4.5 Control variables

Following Aslan et al.'s (2007) method to separate net effect of the firm size on information asymmetry, I include several market and firm accounting variables to control for other influences on the firm's information environment, which might confound my findings.

Since firm size is correlated with liquidity, I include the Amihud measure as a control variable of liquidity. Amihud (2002) develops a measure of illiquidity, defined as a ratio of the absolute value of stock returns to dollar volume, averaged over a particular period. It measures the daily price impact response to one dollar of trading volume. The Amihud measure is also interpreted as a measure of consensus among traders when they receive a new piece of information relevant to stock value. It is given as:

$$\frac{1}{D_q} * \sum_{t=1}^{Q} \frac{|R_t|}{price_t * volume_t} \quad \textbf{(8)}$$

Duarte and Young (2009) also report a positive relation between the Amihud measure and PIN (or AdjPIN).

Aslan et al. (2007) find that the PIN shows some meaningful relations with accounting and market data. They provide a list of variables that are typically used to measure firm characteristics and performance. One of the variables used to represent firm performance is ROA. Profitable firms may be associated with high PINs/ AdjPINs (Aslan et al., 2007). Investors would have greater incentives to seeking out those firms that offer potential returns from information-based trading.

Another proxy of firm trading performance is the standard deviation of stock returns (Lai et al., 2008). This variable measures the uncertainty of a firm, reflecting the firm-level trading environment. Firms with higher volatility may imply more opportunities for informed investors, hence offering more informed trading (French and Roll, 1986). High volatility also means that there is a high cost of gathering and processing accurate information, possibly making institutional investors avoid investing in such firms and, therefore, lowering the arrival rates of informed trading orders. The relation between PIN (or AdjPIN) and stock return volatility may be ambiguous.

Tobin's q is used to measure the firm's investment opportunity. Firms with high Tobin's q may have greater growth options. Similarly, high growth equally means high uncertainty and more opaqueness, thus making the valuation of such firms difficult in information-based trading. Meanwhile, such firms may prove that private information associated with insider trading is likely to be more valuable. This might frighten outside institutional investors, who are usually viewed as being a competitive party to insider trading. As a result, the relation between Tobin's Q and PIN (or AdjPIN) is unclear.

5. Results and Analysis

5.1 Descriptive statistics

The descriptive statistics on all of the variables I have discussed are listed in Table 1. The table provides the cross-sectional mean, median, maximum, minimum, standard deviation, skewness, and kurtosis of the a set of 58 stock samples. The estimated PIN ranges from 0.073 to 0.371 with a mean of 0.242, which is similar to the empirical results of Aslan et al. (2007), based on U.S data. Consistent with Duarte and Young (2009), AdjPIN also similarly ranges from 0.015 to 0.357 with a mean of 0.216. Both the mean and median of AdjPIN are lower than those of for PIN, which is also in line with Duarte and Young (2009)'s finding. The logarithm of the market value of equity (LOGMV) ranges from 7.90 to 13.36 indicating the very diverse firm sizes with the sample. Not surprisingly, the log form of the stock prices(LOGPRICE) range from -0.698 to 5.589 similarly presenting a wide range of stock prices in my sample.

Table1 Descriptive statistics

The PIN is the probability of information-based trading in stock i. The AdjPIN is the adjusted probability of informed trading in stock i. PSOS is the estimated probability of symmetric order flow shock. LOGMV is the logarithm of the market value of equity in firm i at the end of 2005. LOGPRICE is the log form of the yearly-averaged daily stock price. TURNOVER represents the logarithm of averaged daily number of shares traded divided by the number of shares outstanding for firm i in 2006. Amihud is the Amihud measure. ROA is the return on assets. STDEV is the annualised standard deviation of the daily returns for firm i in 2005 and TOBINQ is the Tobin's Q.TURNOVER represents the logarithm of the averaged daily number of shares traded divided by the number of shares outstanding for firm i in 2006. EPSILON is the arrival rate of uninformed trading.

	Mean	Median	Maximum	Minimum	Stdev	Skewness	Kurtosis	Obs
PIN	0.242	0.242	0.371	0.073	0.057	-0.418	4.242	58
ADJPIN	0.216	0.212	0.357	0.015	0.050	-0.409	7.102	58
PSOS	0.281	0.283	0.470	0.137	0.084	0.416	2.490	58
LOGMV	10.086	9.826	13.357	7.895	1.176	0.712	3.352	58
LOGPRICE	2.741	2.779	5.589	-0.698	1.491	-0.252	2.766	58
TURNOVER	-6.500	-6.656	-3.708	-8.453	1.109	0.196	2.306	58
AMIHUD	3.563	1.001	93.426	0.033	12.319	6.879	50.568	58
ROA	0.070	0.073	0.308	-0.456	0.114	-1.778	9.861	58
STDEV	1.686	1.462	5.066	0.249	1.143	0.962	3.351	58
TOBINQ	1.541	1.267	4.377	0.559	0.843	1.952	6.490	58
EPSILON	36.363	26.591	119.322	3.338	24.796	1.070	3.774	58

5.2 Univariate test

Table 2 presents how these variables of PIN and AdjPIN are correlated with other control variables employed in my research. Of particular interest is the positive correlation between firms size (LOGMV) and PIN (and AdjPIN), with correlations of 0.31 and 0.71, respectively. Both of the correlations are positive and significant. This result preliminarily confirms my main finding that large firms in Thailand tend to have a higher intensity of informed trading.

The correlation between PIN and AdjPIN is positive and significant at around 0.41. Both PIN and AdjPIN are negatively correlated with PSOS at around -0.18 and -0.466, respectively. Since PSOS represents illiquidity (Duarte et al, 2009), this result is also in line with my hypothesis that informed trading in the Thai market is negatively correlated with illiquidity. This finding is further proved by the negative correlation which is nearly significant, between the Amihud measure (AMIHUD) and AdjPIN.

AdjPIN and PIN are as expected positively correlated with firm price(LOGPRICE), indicating that higher stock prices can on its own incur a higher intensity of informed trading. This relation in part shows that small investors, who are also usually uninformed investors, may shy away from stocks with high prices, leading to a relative shortage of uninformed investors for large firms. This may contradict finding from the U.S market, but is in line with my hypothesis.

As expected, PIN and the return on assets (ROA) are slightly positively correlated, but AdjPIN is insignificantly negatively correlated with ROA. Firms in Thailand with high investment opportunities high Tobin'q(TOBINQ) are often large firms with high profitability, thus being subject to more informed trading. This implies that the Thai market may also have a pattern of earning concentration, as documented by Dangelo et al. (2004).

A positive correlation between LOGMV and LOGPRICE once again provides the evidence that large firms in Thailand usually have high stock prices. I further rank my sample firms by firm capitalisation to examine whether or not the relation is strictly monotonic. As expected, the size deciles presented in Table 3 clearly show a nearly strict monotonic relation with stock prices

Firm size has an unsurprisingly negative correlation with the proxies of illiquidity, AMIHUD and PSOS. This is consistent with prior studies that find that large firms may have less price impacts, the measures which can also be seen as a proxy of liquidity. The empirical correlation matrix also shows that LOGMV has a positive relation to ROA and TOBINQ, and is negatively related to the annualised standard deviation of the daily returns (STDEV), indicating that larger firms in Thailand are profitable, and of high growth and less volatile in stock prices. In contrary to the relation that large firm stocks have less price impact indicating a positive relation between firm size and liquidity measured by Amihud ratio, turnover (TURNOVER) is negatively related to LOGMV. This indicates that TURNOVER tends to be lower among larger capitalised firms. And it is also not surprising to conclude that firms in Thailand with high stock prices are usually large-cap firms which may be largely invested in by institutional investors. Investments by institutional investors should be larger in scale, and the holding period is longer.

The arrival rate of informed trading, EPSILON, is positively related with LOGMV. In line with Easley et al (2000) and Brown et al (2006), large firms tend to attract more uninformed trading, which is more likely to be originated from individual investors.

LOGPRICE shows a negative relation with PSOS and AMIHUD, indicating that high price stocks in Thailand have more flattened demand curve and less price impact. The positive relations between LOGPRICE and ROA as well as TOBINQ may simply reflect the positive relation between LOGPRICE and LOGMV. Consistent with prior studies, The significantly negative relation between LOGPRICE and TURNOVER indicates that stocks with high price tend to have a lower stock trading activities.

In short, large firms in Thailand, especially in my sample, are often low price impact companies with high profitability, low stock price volatility, and high growth options, while they also have relatively high stock prices. Therefore, large companies are ideal investment targets for institutional investors, particularly for international investors.

Table 2: Correlation Matrix

The PIN is the probability of information-based trading in stock i. The ADJPIN is the adjusted Probability of informed trading in stock i. The estimated probability of the symmetric order flow shock is PSOS. LOGMV is the logarithm of the market value of equity in firm i at the end of 2005. LOGPRICE is the log form of the yearly-averaged daily stock price. AMIHUD is the Amihud measure. ROA is the return on assets. STDEV is the annualised standard deviation of the daily returns for firm i in 2005. TOBINQ is the Tobin's Q. TURNOVER represents the logarithm of averaged daily number of shares traded divided by the number of shares outstanding for firm i in 2006. EPSILON is the arrival rate of uninformed trading.

		AdjPIN	PIN	PSOS	LOGMV	LOGPRICE	AMIHUD	ROA	STDEV	TOBINQ	TURNOVER	EPSILON
AdjPIN		1										
	p-value											
PIN		0.408	1									
	p-value	(0.002)										
PSOS		-0.466	-0.187	1								
	p-value	(0.000)	(0.160)									
LOGMV		0.708	0.311	-0.618	1							
	p-value	(0.000)	(0.017)	(0.000)								
LOGPRICE		0.545	0.503	-0.396	0.68	1						
	p-value	(0.000)	(0.000)	(0.002)	(0.000)							
AMIHUD		-0.201	0.058	0.3	-0.254	-0.059	1					
	p-value	(0.130)	(0.667)	(0.022)	(0.055)	(0.662)						
ROA		-0.018	0.342	-0.387	0.263	0.46	-0.004	1				
	p-value	(0.896)	(0.009)	(0.003)	(0.046)	(0.000)	(0.977)					
STDEV		-0.064	0.023	0.159	-0.146	-0.108	0.094	0.114	1			
	p-value	(0.633)	(0.865)	(0.234)	(0.273)	(0.421)	(0.485)	(0.394)				
TOBINQ		0.266	0.232	0.056	0.244	0.364	-0.058	0.041	-0.101	1		
	p-value	(0.043)	(0.08)	(0.675)	(0.065)	(0.005)	(0.666)	(0.757)	(0.449)			
TURNOVER		-0.059	-0.289	0.077	-0.244	-0.451	-0.183	-0.436	-0.134	-0.155	1	
	p-value	(0.660)	(0.028)	(0.567)	(0.065)	(0.000)	(0.169)	(0.001)	(0.320)	(0.247)		
EPSILON		0.299	-0.155	-0.362	0.328	-0.134	-0.238	-0.201	-0.236	-0.108	0.670	1
	p-value	(0.022)	(0.245)	(0.005)	(0.012)	(0.317)	(0.072)	(0.130)	(0.074)	(0.419)	(0.000)	

Table 3: Relation between LOGMV and LOGPRICE

The table reports the monotonic positive relation between LOGMV and LOGPRICE sorted by size deciles.

DECILE	TICKER	LOGMV	LOGPRICE	average	DECILE	TICKER	LOGMV	LOGPRICE	AVERAGE
HIGH	PTT.BK	13.3569	5.4618	4.3759	6th	MINT.BK	9.7540	2.1392	2.9759
	ADVA.BK	12.6719	4.5161			MAKR.BK	9.7220	4.3205	
	PTTE.BK	12.6415	4.6925			PSL.BK	9.7117	2.8866	
	SCC.BK	12.5872	5.4543			TTA.BK	9.7059	3.0321	
	IRPC.BK	11.9700	1.9922			BECL.BK	9.7051	3.0983	
	TOP.BK	11.7718	4.1385			SATT.BK	9.6960	2.3785	
2th	SHIN.BK	11.7497	3.5644	3.9077	7th	SSI.BK	9.5848	0.1805	1.7221
	PTTC.BK	11.4254	4.4026			CCET.BK	9.5579	1.3625	
	SCCC.BK	11.3386	5.5887			TPC.BK	9.5405	2.7821	
	AOT.BK	11.2252	4.0429			ITV.BK	9.4933	1.7200	
	THAI.BK	11.2219	3.7882			CK.BK	9.4790	2.2904	
	LH.BK	11.1460	2.0595			STEC.BK	9.4415	1.9971	
3th	RATC.BK	10.9929	3.6443	3.3255	8th	GSTE.BK	9.4345	0.1468	1.7818
	CPF.BK	10.7336	1.6734			KSL.BK	9.4255	2.2853	
	EGCO.BK	10.6606	4.4082			VNT.BK	9.3651	2.2037	
	BANP.BK	10.4801	4.9935			STAN.BK	9.3087	5.0379	
	ITD.BK	10.4515	1.8795			TTNT.BK	9.1973	0.9185	
	GLOW.BK	10.4366	3.3542			INOX.BK	9.1860	0.0985	
4th	TRUE.BK	10.4327	2.2467	3.0956	9th	MAJO.BK	9.1369	2.7699	1.3491
	ATC.BK	10.2578	3.3788			ROBI.BK	9.0668	2.3514	
	BGH.BK	10.2246	3.3442			SAMA.BK	8.9619	2.1675	
	BEC.BK	10.2036	2.7766			AP.BK	8.9580	1.3463	
	TUF.BK	10.1854	3.2729			QH.BK	8.8921	0.1569	
	BH.BK	9.9743	3.5547			NSM.BK	8.6179	-0.6977	
5th	HANA.BK	9.9693	3.3078	3.2249	LOW	GOLD.BK	8.3846	2.0412	1.1079
	DELT.BK	9.9593	2.9029			LOXL.BK	8.3129	0.9446	
	UCOM.BK	9.9510	3.8015			PICN.BK	7.9188	-0.6003	
	MCOT.BK	9.8824	3.5403			SOLA.BK	7.8954	2.0460	
	TPIP.BK	9.8509	2.7361						
	RCL.BK	9.8018	3.0608						

5.3 Multivariate test

To further test the robustness of the firm size effect on AdjPIN or PIN, I carefully choose AMIHUD for illiquidity (an inverse measure of liquidity); ROA and STDEV for firms performance; and TOBINQ for the investment opportunity set of a firm. Combined with the control variables, PIN or AdjPIN can be expressed as a function of firm size.

If my finding that large firms in Thailand have high levels of information-based trading is because of liquidity, or is merely due to firm performance or the expectation of growth associated with large firms, I would expect that, when taking into account those control variables, the coefficient of LOGMV would be somewhat insignificant.

Table 4 and Table 5 illustrate that this may not be the case. On the contrary, the coefficient of LOGMV is strongly significant and is positively related to both PIN and AdjPIN in both of the univariate and multivariate regressions. In the Model 8 of Table 4, I obtain a strong significant t-test result (t-statistic=7.33, p- value <0.00001) on the LOGMV coefficient. In Model 8 of Table 5, a coefficient with the same sign is still obtained, indicating a positive relation between PIN and LOGMV, while the t-statistic is marginally significant at 1.7 (and a P-value < 0.094) , which is possibly due to the explanatory power of PIN, as discussed in the previous sections. Among those control variables, only ROA remains significant, but the signs in the two models are different, showing contradictory results. The finding reflects a strong influence of firm size on informed trading of stocks in Thailand. In other words, the trading strategies of informed investors may be designed to put more weight on large and visible firms with little consideration given to other factors,, such as the liquidity and profitability of these firms. After all, the stocks in Thailand are fairly inexpensive relative to stocks in some other countries. Consequently, I can prove my first hypothesis that, **based on the Thai firm samples, large firms have a higher probability of informed trading than do small firms.**

When I examine the Adjusted R-square results, I find that the Model 8 of Table 4 has a high R-squared value of 0.52, which is consistent with the advantage of AdjPIN already discussed. Alternatively, an adjusted R-square of 0.14 is obtained from Model 8 of Table 5. This again might imply that PIN is a noisy measure.

In order to investigate the behaviour of uninformed trading individually, I use the arrival rate of uninformed trading, EPSILON, as a dependent variable for regression on the set of control variables outlined above in Table 6. The EPSILON is in negative and significant relation with LOGPRICE, while remains positive relation with LOGMV. This proves that the activities of uninformed trading decreases as stock prices increase, which is in consistence with the clientele hypothesis. The negative relation is further verified in the multivariate regression, the results of which are shown in Model 7 of Table 6. As a result, I can identify my second hypothesis that **based on a sample set of Thai firms, firms with high prices have a lower level of uninformed trading.**

Table 6 also provides some interesting evidence that large firms attract more informed trading, consistent with Easley et al. (2001) and Brown et al. (2006)'s findings. The significant and positive relation between LOGMV and EPSILON is robust in the multivariate regressions, as suggested in Model 1- 7 of Table 6.

Table 4: Cross-sectional Regression

The independent variable is the adjusted probability of informed trading AdjPIN in stock i. LOGMV is the logarithm of the market value of equity in firm i at the end of 2005. LOGPRICE is a log form of the daily closed price averaged across 2006. TURNOVER represents the logarithm of averaged the daily number of shares traded divided by the number of shares outstanding for firm i in 2006. AMIHUD is the Amihud measure multiplied by 10E+6. STDEV is the annualized standard deviation of the daily returns for firm i in 2005. ROA is the return on assets for firm i in 2005and TOBINQ is the Tobin's Q for firm i in 2005.

MODEL	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	DEPENDENT VARIABLES														
	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN	AdjPIN
INDEPENDENT VARIABLES	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF
INTERCEPT	-0.086	-0.062	-0.063	-0.083	-0.092	-0.104	-0.085	-0.113	0.166	0.225	0.170	0.167	0.162	0.162	0.158
p-value	(0.039)	(0.201)	(0.158)	(0.056)	(0.038)	(0.012)	(0.042)	(0.013)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LOGMV	0.030	0.027	0.031	0.030	0.030	0.032	0.029	0.032							
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)							
LOGPRICE		0.004							0.018	0.022	0.018	0.018	0.023	0.017	0.023
p-value		(0.358)							(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
TURNOVER			0.005							0.011					
p-value			(0.219)							(0.061)					
AMIHUD				-93.700				-53.500			- 686.000				-0.001
p-value				(0.815)				(0.890)			(0.132)				(0.123)
STDEV					0.002			0.004				0.000			0.003
p-value					(0.677)			(0.381)				(0.962)			(0.533)
ROA						-0.096		-0.100					-0.149		-0.151
p-value						(0.024)		(0.021)					(0.006)		(0.007)
TOBINQ							0.006	0.006						0.005	0.002
p-value							(0.311)	(0.298)						(0.522)	(0.826)
OBSERVATUIONS	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58
ADJUSTED R-SQUARED	0.492	0.490	0.497	0.483	0.484	0.529	0.492	0.518	0.285	0.318	0.302	0.272	0.367	0.277	0.364

‘

Table 5 Cross-sectional Regression

The dependent variable is the probability of information-based trading, PIN, in stock i. LOGMV is the logarithm of the market value of equity in firm i at the end of 2005. LOGPRICE is a log form of the daily closed price averaged across year 2006. TURNOVER represents the logarithm of the averaged daily number of shares traded divided by the number of the shares outstanding for firm i in 2006. AMIHUD is the Amihud measure multiplied by 10E+6. STDEV is the annualized standard deviation of daily returns for firm i in 2005. ROA is the return on assets for firm i in 2005 and TOBINQ is the Tobin's Q for firm i in 2005.

MODEL	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	DEPENDENT VARIABLES														
	PIN	PIN	PIN	PIN	PIN	PIN	PIN	PIN	PIN	PIN	PIN	PIN	PIN	PIN	PIN
INDEPENDENT VARIABLES	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF
INTERCEPT	0.091	0.214	0.043	0.071	0.080	0.117	0.094	0.096	0.190	0.168	0.188	0.183	0.192	0.186	0.180
p-value	(0.147)	(0.003)	(0.524)	(0.277)	(0.225)	(0.061)	(0.135)	(0.157)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LOGMV	0.015	-0.003	0.012	0.017	0.015	0.011	0.013	0.011							
p-value	(0.017)	(0.722)	(0.051)	(0.010)	(0.016)	(0.067)	(0.042)	(0.094)							
LOGPRICE		0.021							0.019	0.018	0.019	0.019	0.017	0.018	0.016
p-value		(0.001)							(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.006)
TURNOVER			-0.012							-0.004					
p-value			(0.083)							(0.550)					
AMIHUD				670.000				575.000			401.000				379.000
p-value				(0.270)				(0.331)			(0.455)				(0.487)
STDEV					0.003			0.002				0.004			0.003
p-value					(0.590)			(0.799)				(0.507)			(0.664)
ROA						0.138		0.133					0.069		0.069
p-value						(0.033)		(0.042)					(0.286)		(0.308)
TOBINQ							0.011	0.012						0.004	0.006
p-value							(0.209)	(0.177)						(0.655)	(0.515)

OBSERVATUIONS	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58
ADJUSTED R-SQUARED	0.081	0.227	0.114	0.085	0.069	0.139	0.091	0.138	0.239	0.231	0.233	0.232	0.242	0.228	0.215

Table 6 Cross-sectional Regression

The dependent variable is the arrival rate of uninformed trading, EPSILON, in stock i. LOGMV is the logarithm of the market value of equity in firm i at the end of 2005. LOGPRICE is a log form of the daily closed price averaged across 2006. AMIHUD is the Amihud measure multiplied by 10E+6. STDEV is the annualized standard deviation of the daily returns for firm i in 2005. ROA is the return on assets for firm i in 2005 and TOBINQ is the Tobin's Q for firm i in 2005.

MODEL	1	2	3	4	5	6	7	8	9	10	11	12	13
	DEPENDENT VARIABLES												
	EPSILON	EPSILON	EPSILON	EPSILON	EPSILON	EPSILON	EPSILON	EPSILON	EPSILON	EPSILON	EPSILON	EPSILON	EPSILON
INDEPENDENT VARIABLES	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF	COEF
INTERCEPT	-33.460	-99.316	-23.332	-20.451	-46.085	-34.775	-76.658	42.460	44.892	52.970	41.437	44.428	56.156
p-value	(0.221)	(0.001)	(0.409)	(0.469)	(0.086)	(0.198)	(0.014)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LOGMV	6.923	16.455	6.036	6.330	8.641	7.952	15.044						
p-value	(0.012)	(0.000)	(0.032)	(0.021)	(0.002)	(0.005)	(0.000)						
LOGPRICE		-11.049					-9.279	-2.224	-2.465	-2.678	-0.867	-1.809	-1.038
p-value		(0.000)					(0.002)	(0.317)	(0.258)	(0.221)	(0.727)	(0.451)	(0.694)
AMIHUD			-334.000				-159.000		-497.000				-461.000
p-value			(0.207)				(0.489)		(0.062)				(0.080)
STDEV				-4.168			-3.947			-5.496			-4.697
p-value				(0.133)			(1.109)			(0.056)			(0.104)
ROA					-67.133		-23.440				-38.508		-31.282
p-value					(0.017)		(0.394)				(0.238)		(0.333)
TOBINQ						-5.884	-2.866					-2.015	-3.371
p-value						(0.125)	(0.412)					(0.634)	(0.412)
OBSERVATIONS	58	58	58	58	58	58	58	58	58	58	58	58	58
ADJUSTED R-SQUARED	0.092	0.321	0.102	0.113	0.168	0.114	0.333	0.000	0.045	0.048	0.008	-0.014	0.074

6. Discussion and conclusion

In my research, I examine the relation between informed trading and firm size using a sample of Thai firms for 2006. Although most of the previous studies find that large firms usually bear a low probability of information-based trading, I find significant and robust evidence of an opposite empirical pattern. I hereby provide some possible explanations on these phenomena anomaly.

According to the findings of many prior studies, it is the visibility of a firm that may attract more outside informed traders, especially international institutional investors. The investment preference of large firms may be partially explained by some institutional investors' constraints such as considerations of liquidity risk, prudent man rule and so forth. This relation is justified by Dennis and Weston (2001). As a result, the institutional holdings of larger Thai firms may be higher than smaller firms. The informed trading may be reflected in order flow imbalance of large firms through such mechanisms as 'contagion' across markets and 'cross-market trading'.

The mechanism of 'contagion' across both the two markets (Main board and Alien board) (Bailey and Jagtiani, 2006) is a multiple markets trading behaviour where investors in the two markets may observe the stock price trend in one market and simultaneously make the trend spread (contagion) over and arise in the same direction in the other market. The mechanism of 'cross-market trading' (Bailey et al., 2008) refers to that foreign investors may also have a cross-market trading channel to capitalise their value-enhancing information, because their trading accounts for 15 % of the Main board trading, and is almost certainly informed trading. These mechanisms may constitute the reason for the relatively high level of informed trading in large cap Thai firms. In light of the definition of PIN and AdjPIN, I label this effect the ***Numerator Effect*** inspired by DCF valuation terminology.

On the other hand, larger Thai firms tend to have higher stock prices and larger ticker size. Some researches (Pavabutra and Sirodom, 2007, Pavabutra and Prangwattananon 2009); which study stock spilt impacts on the liquidity of the Thai Market indicate that there is a distinct effect of optimal trading range from THB 10 to THB 25. This *clientele* *hypothesis* implies that lower stock prices attract more individual investors, who are usually regarded as uninformed investors. In fact, it is only the low price stocks that individual investors, who dominate the trading activities on the Thai stock market, particularly prefer to invest in. Pavabutra and Prangwattananon (2009) provide evidences that the number of individual investors decreases monotonically as stock prices increase. The reason may be associated with consideration of trading costs. In my sample, the stock prices ranges from 1 to more than THB 400, and nearly one third of the companies are above the upper bound of optimal trading range of THB 25. This deviation from optimal trading range may lead to a higher intensity of informed trading for larger firms in my sample. The possible different trader composition across firms of different sizes in the Thailand market further assists in this verification, and arriving at one of the contributions of this thesis, which is the confirmation that the stock price effect is a proxy of the preference of uninformed investors, and may matter to informed trading in a firm. The thesis also provides implications from a market microstructure perspective that, in terms of the preferred trading range hypothesis, large Thai firms with relatively high stock prices should implement a stock split policy to attract more public investors and widen the clientele of their firms. This will be of benefit in liquidity provision.

Moreover, to avoid losing to the recognised informed investors in large Thai firms, such as international investors, uninformed investors may reduce the proportion of their investment in stocks with high levels of informed trading and higher institutional holdings. Consequently, this adverse selection effect may be attributed to a low level of uninformed trading. Similarly, I refer to this effect as the ***Denominator Effect***. This is an empirical matter of trader composition. Adding these two aspects together may arrive at a high intensity of informed trading, as measured by PIN and AdjPIN in large Thai firms.

Several papers (Bardong et al., 2007; Aslan et al., 2007; Duarte and Young, 2009) suggest that using a full year of intraday trade data will better achieve reliable values of PIN and AdjPIN. This estimation horizon, however, may tend to capture only long term effects of informed trading which is likely to come from the strategic cost-minimising behaviour of informed traders (Kyle 1985). In the meantime, it may also rule out short-term informed trading, which is more significant, especially for small firms, which leads to a model bias. All of these findings may imply that other measures of informed trading that can be applied to a short horizon should be appropriate when implemented together.

References

1. Admati, A. R. and Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies*, 1, 3-40.

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

3. Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5, 31-56.

4. Aslan, H., Easley, D., Hvidkjaer, S. and O'Hara, M. (2007). *Firm Characteristics and Informed Trading: Implications for Asset Pricing.* Cornell University Working Paper.

5. Bailey, W. and Jagtiani, J. (1994). Foreign ownership restrictions and stock prices in the Thai capital market. *Journal of Financial Economics*, 36, 57-87.

6. Bailey, W., Mao, C. and Sirodom, K. (2006). *Local, foreigners, and multi-market trading of equities: Some intraday evidence.* Cornell University Working Paper.

7. Bailey, W., Mao, C. and Sirodom, K. (2008). *Local, foreigners, and multi-market trading of equities: Some intraday evidence.* Cornell University Working Paper.

8. Baker, H. and Gallagher, P. L. (1980). Management's View of Stock Splits. *Financial Management*, 9, 73-77.

9. Banz, R.W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics,* 9, 3–18.

10. Barclay, M. J. and Warner, J. B. (1993). Stealth Trading and Volatility: Which Trades Move Prices? *Journal of Financial Economics*, 34:3, 281-306.

11. Bardong, F., Söhnke M. B. and Pradeep K. Y. (2007). *Informed Trading, Information Asymmetry, and Pricing of Information Risk: Empirical Evidence from the NYSE.* Lancaster University and University of Oklahoma Working Paper,

12. Barry, C. and Brown, S. (1985). Differential information and security market equilibrium. *Journal of Financial and Quantitative Analysis*, 20 (December), 407-422.

13. Benos, E. and Jochec, M. (2007). *Testing the PIN variable.* University of Illinois at Urbana Champaign Working Paper.

14. Bernhardt, D. and Hughson, E. (1997). Splitting Orders. *Review of Financial Studies*, 10:1, 69-101.

15. Brown, S., Hillegeist, S. A. and Lo, K. (2004). Conference Calls and Information Asymmetry. *Journal of Accounting and Economics*, 37, 343-366.

16. Brown, S., Hillegeist, S. and Lo, K. (2006). *The Effect of Meeting or Missing Earnings Expectations of Information Asymmetry.* University of British Columbia Working Paper.

17. Chan, K., Menkveld, A. J. and Yang, Z. (2008). Information Asymmetry and Asset Prices: Evidence from the China Foreign Share Discount. *Journal of Finance*, 63(1), 159–196.

18. Chakravarty, S. (2001). Stealth-trading: Which Trader's Trades Move Stock Prices? *Journal of Financial Economics*, 61:2, 289-307.

19. Choe, H., Kho, B. and Stulz, R. (1999). Do foreign investors destabilize stock Markets? The Korean experience in 1997. *Journal of Financial Economics*, 54, 227-264.

20. Chordia, T., Huh, S. W. and Subrahmanyam, A. (2006). The Cross-Section of Expected Trading Activity. forthcoming in *Review of Financial Studies*.

21. Chordia, T. and Subrahmanyam, A. (2004). Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Economics*, 72, 485-518.

22. Ciner, C. and Karagozoglu, A. K. (2008). Information asymmetry, speculation and foreign trading activity: Emerging market evidence. *International Review of Financial Analysis*, 17, 664–680.

23. Claessens, S., Djankov, S. and Lang, L. (2000). The separation of ownership and control in East Asian corporations. *Journal of Financial Economics*, 58, 8, 112.

24. Copeland, T. E. (1979). Liquidity changes following stock splits. *Journal of Finance*, 34, 115-141.

25. Covrig, V., Lau, S. T. and Ng, L. (2006). Do domestic and foreign fund managers have similar preferences for stock characteristics? A cross-country analysis. *Journal of International Business Studies*, 37, 407-429.

26. Dahlquist, M. and Goran R. 2003, Direct foreign ownership, institutional investors, and firm characteristics, Journal of Financial Economics 59, 413-440.

27. Dennis, P. J. and Weston, J. P. (2001). *Who's informed? An analysis of stock ownership and informed trading.* Rice University Working Paper.

28. De Angelo, H., De Angelo, L. and Skinner, D. J. (2004). Are Dividends Disappearing? Dividend Concentration and the Consolidation of Earnings. *Journal of Financial Economics*, 72, 425-456.

29. Duarte, J. and Young, L. (2009). Why is PIN priced? *Journal of Financial Economics,* 91(2), 119-138.

30. Easley, D., Engle, R. F., O'Hara, M. and Wu, L. (2008). Time-varying arrival rates of informed and uninformed trades. *Journal of Financial Econometrics,* 6(2), 171-208.

31. Easley, D., Hvidkjaer, S. and O'Hara, M. (2002). Is information risk a determinant of asset returns? *Journal of Finance*, 57, 2185-2222.

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

33. Easley, D. and O'Hara, M. (1992). Adverse selection and large trade volume: The implications for market efficiency. *Journal of Financial and Quantitative Analysis*, 27, 185-208.

34. Easley, D., O'Hara, M. and Saar, G. (2001). How stock splits affect trading: A microstructure approach. *Journal of Financial and Quantitative Analysis*, 36, 25-51.

35. Falkenstein, E. G. (1996). Preferences for stock characteristics as revealed by mutual fund holdings. *Journal of Finance,* 51, 111-135.

36. Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics,* 33, 3-56.

37. Ferreira, M. A. and Matos, P. P. (2007). The Colour of Investors' Money: The Role of Institutional Investors Around the World. Forthcoming, *Journal of Financial Economics.*

38. French, K. R. and Roll, R. (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics*, 17, 5-26.

39. Gompers, P. and Metrick, A. (2001). Institutional investors and equity prices. *The Quarterly Journal of Economics*, 116, 229-259.

40. Hasbrouck, J. (1991). Measuring the Information Content of Stock Trades. *Journal of Finance*, 46(1), 179-207.

41. Hasbrouck, J. (1991a). The Summary Informativeness of Stock Trades: An Econometric Analysis. *Review of Financial Studies*, 4(3), 571-595.

42. Hasbrouck, J. (2007). *Empirical Market Microstructure.* Oxford: Oxford University Press.

43. Horowitz, J., Loughran, T. and Savin, N. (2000). Three analyses of the firm size premium. *Journal of Empirical Finance*. vol. 7, 143-53.

44. Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53, 1315–1335.

45. Lakonishok, J. and Lev, B. (1987). Stock Splits and Stock Dividends: Why, Who and When. *Journal of Finance,*42, 913-932.

46. Lai, S., Ng, L. and Zhang, B. (2008). *Informed Trading Around The World.* Retrieved on [date here] from SSRN: http://ssrn.com/abstract=1319668

47. Lee, C. and Ready, M. (1991). Inferring trade direction from intraday data. *Journal of Finance*, 46, 733-746.

48. Merton, R. C. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *Journal of Finance*, 42, 483-510.

49. Pavabutr, P. (2008). The Impact of Stock Splits on Price and Liquidity on the Stock Exchange of Thailand. *International Research Journal of Finance and Economics,* Issue 20.

50. Pavabutr, P. and Sirodom, K. (2007) Stock Splits in a Retail Dominant Order Driven Market, Thammasat University Working Paper .

51. Pavabutra, P. and Prangwattananon, S. (2009). Tick Size Change on the Stock Exchange of Thailand. *Review of Quantitative Finance and Accounting*, 4, 351-371 .

52. Reinganum, M. R, (1981). Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings' Yields and Market Values. *Journal of Financial Economics,* 9, 19-46.

53. Vega, C. (2005). Stock Price Reaction to Public and Private Information. *Journal of Financial Economics,* 82(1), 103–133.

54. Venter, J. H. and de Jongh, D. (2004). Extending the EKOP model to estimate the probability of informed trading. In: *Proceedings of First African Finance Conference*, Cape Town, South Africa

55. Weld, W. C., Michaely, R., Thaler, R. H. and Benartzi, S. (2009). The Nominal Share Price Puzzle. *Journal of Economic Perspectives,* Vol. 23 Issue 2, 121-142.

Appendix A: Raw Data

ticker	PSOS	AdjPIN	PIN	ROA	STDEV	LOGMV	Amihud	Turnover	TOBINQ	Logprice
ADVA.BK	0.2019	0.2572	0.2721	0.1305	0.4469	12.6719	0.0873	-7.1119	2.9357	4.5161
AOT.BK	0.3053	0.2526	0.2771	0.0690	0.5126	11.2252	0.1357	-6.7109	1.0261	4.0429
AP.BK	0.3862	0.1586	0.2294	0.1290	2.9641	8.9580	4.3994	-6.2253	1.2611	1.3463
ATC.BK	0.2395	0.2174	0.2570	0.1145	1.9510	10.2578	1.4515	-5.8668	1.1002	3.3788
BANP.BK	0.1856	0.2570	0.3282	0.0714	1.4668	10.4801	0.1405	-5.8272	1.2381	4.9935
BEC.BK	0.2809	0.2053	0.2666	0.2273	2.6987	10.2036	12.6794	-7.0513	3.8793	2.7766
BECL.BK	0.1788	0.2022	0.2365	0.0337	0.2487	9.7051	0.3692	-6.2789	1.0191	3.0983
BGH.BK	0.2519	0.2164	0.2278	0.0525	0.3451	10.2246	1.1144	-7.6440	1.6909	3.3442
BH.BK	0.3078	0.2124	0.2231	0.1654	0.3328	9.9743	1.0226	-7.4251	3.7765	3.5547
CCET.BK	0.2836	0.1777	0.2255	0.0725	2.3104	9.5579	4.4318	-7.7826	1.0465	1.3625
CK.BK	0.3062	0.1841	0.1849	-0.0400	1.1429	9.4790	0.7576	-5.8021	1.2731	2.2904
CPF.BK	0.1881	0.2459	0.2546	0.0263	0.4964	10.7336	0.1674	-6.2563	1.0218	1.6734
DELT.BK	0.2817	0.1768	0.2185	0.0749	1.3364	9.9593	1.4208	-6.9696	1.2505	2.9029
EGCO.BK	0.2002	0.1906	0.2535	0.0840	1.6037	10.6606	0.5793	-7.0272	1.0843	4.4082
GLOW.BK	0.2750	0.2466	0.2725	0.1025	1.4834	10.4366	0.8379	-7.2435	1.1400	3.3542
GOLD.BK	0.4231	0.1536	0.2598	0.0215	1.6757	8.3846	93.4260	-7.1774	0.8771	2.0412
GSTE.BK	0.2431	0.2019	0.2169	0.0398	1.8182	9.4345	0.2385	-4.9825	0.6620	0.1468
HANA.BK	0.2376	0.2305	0.2577	0.1795	3.4411	9.9693	2.2341	-6.7343	1.9305	3.3078
INOX.BK	0.2889	0.1774	0.2154	0.1099	3.2294	9.1860	3.4447	-7.5891	0.8617	0.0985
IRPC.BK	0.1734	0.2415	0.2485	0.0533	1.8936	11.9700	0.0401	-5.8917	1.5855	1.9922
ITD.BK	0.2360	0.2316	0.2011	-0.0423	0.7459	10.4515	0.2096	-5.4470	1.3870	1.8795
ITV.BK	0.4478	0.2598	0.2176	-0.4558	1.2143	9.4933	0.9716	-4.9432	4.2515	1.7200
KSL.BK	0.3564	0.2080	0.2533	0.0626	2.0073	9.4255	3.2982	-7.2168	1.4501	2.2853
LH.BK	0.1449	0.2519	0.1719	0.0778	0.4717	11.1460	0.6549	-6.3560	2.0973	2.0595
LOXL.BK	0.2954	0.1567	0.2046	0.0686	1.3435	8.3129	1.1768	-4.8212	0.9723	0.9446
MAJO.BK	0.2349	0.1986	0.2058	0.0792	1.9955	9.1369	0.9802	-6.1159	1.5802	2.7699
MAKR.BK	0.4516	0.2116	0.2492	0.0634	1.5700	9.7220	7.6846	-8.2345	1.4117	4.3205
MCOT.BK	0.4148	0.2054	0.3074	0.1760	1.3441	9.8824	1.3543	-6.8508	2.4953	3.5403
MINT.BK	0.3614	0.2085	0.2675	0.0700	1.4743	9.7540	3.2579	-7.1575	1.4798	2.1392
NSM.BK	0.4328	0.1748	0.1061	-0.1115	3.0267	8.6179	1.3819	-4.8770	0.5615	-0.6977
PICN.BK	0.4701	0.1867	0.0985	-0.2236	1.5227	7.9188	1.3089	-3.7077	1.2109	-0.6003
PSL.BK	0.2231	0.2222	0.2626	0.2912	3.8669	9.7117	2.1401	-5.6455	1.3505	2.8866
PTT.BK	0.1573	0.3280	0.0733	0.1270	2.2601	13.3569	0.0370	-6.7162	1.3704	5.4618
PTTC.BK	0.2340	0.2150	0.2672	0.1406	1.0006	11.4254	0.1912	-7.3008	1.0117	4.4026
PTTE.BK	0.2587	0.3006	0.3265	0.1777	0.4533	12.6415	0.0331	-4.9982	2.3979	4.6925
QH.BK	0.2898	0.1677	0.2091	0.0492	4.5360	8.8921	2.7430	-6.3865	0.9364	0.1569
RATC.BK	0.2133	0.2623	0.2900	0.0847	3.6818	10.9929	0.7519	-7.2783	1.3433	3.6443
RCL.BK	0.2841	0.1939	0.2066	0.1179	0.5738	9.8018	1.4337	-6.2913	1.1276	3.0608
ROBI.BK	0.3150	0.0149	0.1839	0.3083	1.0544	9.0668	2.4384	-7.2567	1.4046	2.3514
SAMA.BK	0.2996	0.2249	0.3153	0.1395	0.4130	8.9619	0.4011	-5.4074	1.1914	2.1675

SATT.BK	0.2670	0.1870	0.1824	-0.0014	0.7265	9.6960	0.5225	-5.7376	1.0827	2.3785
SCC.BK	0.1372	0.3155	0.3707	0.1359	3.6707	12.5872	0.0891	-6.9452	1.9454	5.4543
SCCC.BK	0.1967	0.2601	0.2737	0.1905	0.5273	11.3386	0.9062	-7.9160	4.3772	5.5887
SHIN.BK	0.3345	0.2142	0.3315	0.0407	1.9119	11.7497	2.1174	-8.4527	1.9114	3.5644
SOLA.BK	0.3868	0.2012	0.3148	0.0404	1.2192	7.8954	2.8850	-5.3123	2.3566	2.0460
SSI.BK	0.2941	0.1867	0.2210	0.0603	1.4578	9.5848	3.6317	-6.6007	0.8771	0.1805
STAN.BK	0.3850	0.2252	0.2833	0.1921	5.0664	9.3087	13.9865	-8.3183	2.2066	5.0379
STEC.BK	0.3053	0.1912	0.2358	-0.1126	2.7844	9.4415	0.6238	-4.7555	1.5487	1.9971
THAI.BK	0.2199	0.2826	0.2891	0.0367	2.3192	11.2219	0.4761	-7.0900	1.0370	3.7882
TOP.BK	0.1488	0.3568	0.3658	0.1328	0.5001	11.7718	0.0414	-5.9146	1.4496	4.1385
TPC.BK	0.1869	0.2049	0.2082	0.1123	1.2359	9.5405	2.0487	-7.8585	1.0863	2.7821
TPIP.BK	0.2929	0.1822	0.2114	0.0406	0.5664	9.8509	0.3589	-5.5102	0.5593	2.7361
TRUE.BK	0.2934	0.2221	0.2199	-0.0432	0.6105	10.4327	0.1996	-5.6313	1.2350	2.2467
TTA.BK	0.2173	0.2074	0.2296	0.1514	1.4303	9.7059	0.8839	-4.8095	1.1897	3.0321
TTNT.BK	0.3571	0.1633	0.2055	-0.0308	2.1461	9.1973	0.5365	-5.1509	0.8872	0.9185
TUF.BK	0.3952	0.2220	0.2568	0.0730	3.5084	10.1854	2.8859	-7.8358	1.4586	3.2729
UCOM.BK	0.3050	0.2361	0.2347	0.0980	0.8669	9.9510	6.3775	-8.2179	1.5122	3.8015
VNT.BK	0.2231	0.2556	0.2784	0.0299	1.2835	9.3651	6.6771	-8.3429	0.9667	2.2037

Appendix B: SAS codes

```
/*programme 1, Separating original data file */

%macro company_separation (filenumber,filename);

%let count=1;

  %do  number1=1 % to &filenumber;

data data.sample;
infile "C:\Documents and
Settings\LabUser\Desktop\autoseperation20090912csv\&filename&number1..
csv" firstobs=2 missover dsd;

informat
RIC $9.
date  date10.  /* $10. ddmmyy10.*/
Time time10. /* $10.*/
M 4.
type $5.
price best12.
Bidprice best12.
askprice best12.
QUALIFIERS $29.
;


input RIC $ DATE $ Time $ M type $  PRICE  BIDPRICE  Askprice
QUALIFIERS $@;

format
RIC $9.
date  ddmmyy10.  /* $10.date10. */
Time time10./* $10.*/
M 4.
type $5.
price best12.
Bidprice best12.
askprice best12.
QUALIFIERS $29.

;
run;

proc sort data=data.sample;
by ric;
run;

data data.sample1;
set data.sample;
by ric;
if first.ric=1 then do;
companynum+1;
output data.sample1 ;
end;
else do
companynum=companynum;
output data.sample1;
```

```
end;
run;

data data.samplecompanycount;
set data.sample1;
by companynum;
if first.companynum=1;
run;


proc append base = data.totalcompany data = data.samplecompanycount;
run;


data _null_;
set data.samplecompanycount end=last;
if last then do;
call symputx('countnumber',_n_);

end;
run;

%let count1=%eval(&count+&countnumber-1);
%put count1=&count1 count=&count countnumber=&countnumber;
run;

%let x=1;
%do  number=&count % to &count1 ;

data data.companyname&number( keep= ric Date Time TYPE Price Bidprice
Askprice Qualifiers ) ;
set data.sample1;
where companynum= &X;
run;

data data.sample1;
set data.sample1;
if companynum= &x then delete;
run;

%let x=%eval(&x+1);
%put x=&&x;

%end;

%let count= %eval(&count1+1);
%put step2count=&count;

%end  ;

%mend ;


data data.totalcompany;
set data.samplecompanycount;
delete;
run;
```

```
%company_separation (16,data);

data data.totalcompany;
set data.totalcompany;
number=_n_;
run;

proc sort data= data.totalcompany out= data.totalcompany_sort;
by ric;
run;

Run;
```

/*programme 2, processing firm trading and accounting data*/

```
libname sample 'E:\sas\20091009\cleaned firm characteristic data';
%macro readxlsdata (address,file);

LIBNAME xlslib "&address&file..xls";
DATA sample.&file;
SET xlslib.'1$'n;
RUN;

proc sort data=sample.&file out=sample.&file;
by ticker;
quit;
%mend;


%macro trantolong(address,file,va);

LIBNAME xlslib "&address&file..xls";
DATA sample.&file;
SET xlslib.'1$'n;
RUN;

PROC TRANSPOSE DATA =sample.&file
OUT = sample.&file(rename = (Col1 = &va _LABEL_=ticker) drop = _name_);
BY date;
RUN;


proc sort data=sample.&file out= sample.&file;
by ticker;
label ticker = 'ticker';
quit;


data sample.&file;
set sample.&file;
x = Index(ticker, '#');
substr(ticker,x,1)='.';
drop x;
run;

%mend;


/* YEARLYMV MACRO*/
%macro yearlymv(file);

data sample.yearlyMV(drop=MV);
set sample.&file;   *MV20091008;
LOGMV= log(MV);
run;

%mend;
```

```
/* yearlyAmihud*/
%macro yearlyAmihud(dailyprice,va);

data sample.Amihud_daily;
merge sample.&dailyprice sample.&va;
by ticker;
run;

data sample.Amihud_daily;
set sample.Amihud_daily;
where dailyprice is not missing ;
where same and va is not missing;
run;

data sample.amihud_daily;
set sample.amihud_daily;
dailyreturn= Abs(log(dailyprice)-log(lag(dailyprice)));
If dailyreturn=. then delete;
amihud= dailyreturn/va ;
run;

proc means data= sample.amihud_daily noprint MEAN ;
output out = sample.yearlyamihud;
var amihud;
by ticker;
run;

data sample.yearlyamihud(drop=_TYPE_ _STAT_ _FREQ_);
set sample.yearlyamihud;
If _STAT_='MEAN';
run;

%mend;




/*ROA MACRO*/
%macro yearlyROA(Netincome,Totalasset);

data sample.ROA;
merge sample.&Netincome sample.&Totalasset;
by ticker;
run;


data sample.yearlyROA(drop= NI TA);
set sample.ROA;
ROA= NI/TA;
run;

%mend;



%macro yearlyturnover(VO,NOSH);

data sample.dailyturnover;
merge sample.&VO sample.&nosh;
by ticker;
run;
```

```
data sample.dailyturnover;
set sample.dailyturnover;
where vo is not missing;
run;

data sample.dailyturnover;
set  sample.dailyturnover;
turnover=log(vo/nosh);
run;

proc means data= sample.dailyturnover noprint MEAN ;
output out = sample.yearlyturnover;
var turnover;
by ticker;
run;

data sample.yearlyturnover (drop=_TYPE_ _STAT_ _FREQ_);
set sample.yearlyturnover;
If _STAT_='MEAN';
run;

%mend;

%macro yearlytobinq(MV,liabilities,Totalasset);

data sample.TOBINQ;
merge sample.&MV sample.&liabilities sample.&Totalasset ;
by ticker;
run;


data sample.yearlyTOBINQ(drop= MV DEBT TA );
set sample.TOBINQ;
TOBINQ= (MV*1000+DEBT)/TA;
run;

%mend;

%macro yearlySTD(dailyprice);
data sample.dailystd;
set sample.&dailyprice ;
where dailyprice is not missing ;
run;

data sample.dailystd;
set sample.dailystd;
dailyreturn= log(dailyprice)-log(lag(dailyprice));
If dailyreturn=. then delete;
run;

proc means data= sample.dailystd noprint STD ;
output out = sample.yearlystd;
var dailyreturn;
by ticker;
run;

data sample.yearlystd(drop=_TYPE_ _STAT_ _FREQ_ );
set sample.yearlystd;
If _STAT_='STD';
run;
```

```
data sample.yearlystd (drop=dailyreturn);
set sample.yearlystd;
STDEV = sqrt(255)*dailyreturn;
run;

%mend;

%macro yearlyaverageprice(dailyprice);

data sample.dailyprice;
set sample.&dailyprice ;
where dailyprice is not missing ;
run;



proc means data= sample.dailyprice noprint mean ;
output out = sample.yearlyaverageprice;
var dailyprice;
by ticker;
run;



data sample.yearlyaverageprice(drop=_TYPE_ _STAT_ _FREQ_ dailyprice);
set sample.yearlyaverageprice;
If _STAT_='MEAN';
logprice=log(dailyprice);
run;

%mend;

%macro combineyearlydata;

data sample.companyyearlydata;
attrib ticker length= $40;
merge sample.yearlymv sample.yearlyamihud sample.yearlyroa
sample.yearlyturnover sample.yearlytobinq  sample.yearlystd
sample.yearlyaverageprice ;
by ticker;
run;
%mend;



%trantolong(C:\Documents and Settings\LabUser\Desktop\firm
characteristic 1008\, dailyprice20091008,dailyprice);
%trantolong(C:\Documents and Settings\LabUser\Desktop\firm
characteristic 1008\, vo20091008,vo);
%trantolong(C:\Documents and Settings\LabUser\Desktop\firm
characteristic 1008\, va20091008,va);

%readxlsdata(C:\Documents and Settings\LabUser\Desktop\firm
characteristic 1008\, mv20091008);
%readxlsdata(C:\Documents and Settings\LabUser\Desktop\firm
characteristic 1008\, liabilities20091008);
%readxlsdata(C:\Documents and Settings\LabUser\Desktop\firm
characteristic 1008\, netincome20091008);
%readxlsdata(C:\Documents and Settings\LabUser\Desktop\firm
characteristic 1008\, nosh20091008);
```

```
%readxlsdata(C:\Documents and Settings\LabUser\Desktop\firm
characteristic 1008\, totalasset20091008);
%readxlsdata(C:\Documents and Settings\LabUser\Desktop\firm
characteristic 1008\, MTBV20091008);

%yearlymv(mv20091008);
%yearlyAmihud(dailyprice20091008,va20091008);
%yearlyroa(Netincome20091008, Totalasset20091008);
%yearlyturnover(Vo20091008, nosh20091008);
%yearlytobinq(MV20091008,liabilities20091008, Totalasset20091008);
%yearlySTD(dailyprice20091008);
%yearlyaverageprice(dailyprice20091008);
%combineyearlydata
run;
```

/*programme 3, summarizing numbers of daily buy and sell, computing PIN and Adjpin*/

```
options nodate nonotes nosource;
ODS NORESULTS;

/*csv  sample */
libname sample 'C:\Documents and Settings\LabUser\Desktop\test
qualifer 20090915';
run;

proc printto print="C:\Documents and Settings\LabUser\Desktop\test
qualifer 20090915\output.TXT";
run;
proc printto log ="C:\Documents and Settings\LabUser\Desktop\test
qualifer 20090915\log.txt";
run;

/*sample*/
/*trades */
data sample.trades_sum;
set  sample.trades_sum;
delete;
run;

/*PIN*/
data sample.pin_sum;
set  sample.pin_sum;
delete;
run;

/* pin*/
data sample.pin_sum_end;
set  sample.pin_sum_end;
delete;
run;

/*adjpin*/
data sample.adjpin_sum;
set  sample.adjpin_sum;
delete;
run;

/**/
data sample.adjpin_sum_end;
set  sample.adjpin_sum_end;
delete;
run;

/* */
%macro sample;

data sample.daycount;
set sample.E;
by num;
if first.num=1;
run;


data _null_;
set sample.daycount end=last;
```

```
if last then do;
call symputx('daynumber',_n_);
end;
run;

  %do k=1 % to &daynumber;
Data  sample.F;
          set sample.E;
          where num = &k;
run;
/*   *10*/

data sample.G;
set sample.F;
time_record= HMS(hour,minute, second);
run;

/*,*/
data G;
set sample.G;
where time_record < 34200;
run;

data sample.H;
do i=1 to end;
set G point=i nobs=end;
if QUALIFIERS='Open[USER]' then  do k=i+1 to end;
set G point=k;
output;
end;
end;
stop;
run;

/*trade quote*/
data sample.trade(drop=bidprice askprice) sample.quote(drop=price);
set sample.H;
if type='Trade' then output sample.trade;
if type='Quote' then output sample.quote;
run;

/*bid ask*/
data sample.bid (drop=askprice) sample.ask(drop=bidprice);
set sample.quote;
run;

/* timebid ask*/

proc sort data=sample.bid out=bid_order;
by RIC date time;
run;

proc sort data=sample.ask out=ask_order;
by RIC date time;
run;

data ask_order_new;
set ask_order;
if missing(askprice) then delete;
run;
```

```
data bid_order_new;
set bid_order;
if missing(bidprice) then delete;
run;

data quote_new;
merge bid_order_new(in=a) ask_order_new(in=b);
bid = a;
ask = b;
by ric date time;
run;


/*quote_new*/

proc sort data=quote_new;
by ric date time_record ;
run;

data quote_new (drop= bid ask);
set quote_new ;
by ric date time_record;
retain _bidprice _askprice;
if missing(bidprice) then bidprice = _bidprice; else _bidprice =
bidprice;
if missing(askprice) then askprice = _askprice; else _askprice =
askprice;
drop _bidprice _askprice ;
run;

/*bid>askbid ask  sample.quote_new */
data sample.quote_new;
set quote_new;
if bidprice>askprice or bidprice < 0 or askprice <0 then do;
delete;
put ric= sum= bidprice= askprice=;
end;
run;

/*bid ask*/


data sample.quote_new;
set sample.quote_new;
spread = askprice-bidprice;
midpoint= (askprice+bidprice)/2;
if spread > midpoint*0.25 then do;
delete;
put ric= sum= spread= midpoint=;
end;
run;


/*5trade*/
Data r0;
Set sample.trade;
timedifference = time_record -lag(time_record);
Run;


data R6;
```

```
set r0;
di=_n_;
if di=1 then do;
cumsum=5; /* this code ensures that we always keep the first
observation*/
end;
if di>1 then do;
retain cumsum;
if cumsum>=5 then cumsum=timedifference; else
cumsum=timedifference+cumsum;
end;
run;


data sample.trade_merge5second (drop = di timedifference cumsum);
set r6;
where cumsum > 5;
run;



/* trade5*/
data sample.trade_timeadjust;
set sample.trade_merge5second;
time_real= time_record;
time_record= time_record-5;
run;



/* trade */
data sample.trade_new;
set sample.trade_timeadjust;
if _n_ =1 then do;
k= price ;
lagprice = price;
end;
retain k;
retain lagprice;
if    price < k*0.5 or price > k*1.5 then do;
if price< lagprice*0.5 or price>lagprice*1.5 then do;
put ric= sum= price= k=;
delete;
end;
end;
else do;
output;
lagprice = k;
k = price;
end;
run;

/**/
data sample.quote_one;
set sample.quote_new;
keep  Ric date  time_record type bidprice askprice sum;
run;

data sample.trade_one;
set sample.trade_new;
keep Ric date time_record type price sum;
```

```
run;

proc sort data=sample.quote_one out = p2;
by RIC date time_record ;
run;

proc sort data=sample.trade_one out=p3;
by RIC date time_record ;
run;

data sample.ticker (drop = bid ask);
merge p2(in=a) p3(in=b);
bid = a;
ask = b;
if a then ticker = 1; else ticker= 0;
*if bid=0 or ask=0 then output;
by ric date time_record;
run;

/*QUOTE */
data Q1;
set sample.ticker;
run;


proc sort data=q1;
      by ric date time_record type;
      run;

data q2;
set q1;
by ric date;
retain _bidprice _askprice;
if missing(bidprice) then bidprice = _bidprice; else _bidprice =
bidprice;
if missing(askprice) then askprice = _askprice; else _askprice =
askprice;
drop _bidprice _askprice;
run;


/* quote */

data q3;
set q2;
midpoint = (bidprice+askprice)/2;
if type='Quote' then delete;
run;

data q31;
set q3;
lagmidpoint = lag(midpoint);
lag2midpoint= lag2(midpoint);
if price > midpoint then buy = 1;
if price < midpoint then sell = -1;
if price = midpoint then do;
if price > lagmidpoint and lagmidpoint ^= .  then buy =1;
if price < lagmidpoint and lagmidpoint ^= . then sell = -1;
if  price =lagmidpoint and lagmidpoint ^= . then do;
if price > lag2midpoint and lag2midpoint ^= . then buy =1;
if price < lag2midpoint and lag2midpoint ^= . then sell =-1;
```

```
if price = lag2midpoint  then do;
buysell = 0;
put ric= sum= buysell=;
end;
end;
end;
run;

      /**/

proc sort data=q31;
by ric date time_record;
run;


proc freq data = q31 ;
table buy /out=temp1;
table sell/out = temp2;
by ric date;
run;

data Q4;
set temp1 temp2;
if sell ^= . then do;
sells = COUNT;
output;
end;
if buy ^= . then do;
buys= count;
output;
end;
run;


data sample.onedaysellbuy( keep = ticker date buys sells );
set q4;
rename ric = ticker;
buys = lag(buys);
if sell = -1 then output;
run;

proc append base = sample.trades data = sample.Onedaysellbuy;
run;

%end;

%mend sample;



%macro  company (companynumber , name);
%local number;
%do  number=1 % to &companynumber;


data sample.sample;
SET Sample.&name&number;
run;

/* sample.sample  price nid ask  */
data sample.a;
```

```
set sample.sample;
where  PRICe or BIDPRICE or Askprice is not missing;
run;

/* sample.b price */
data sample.b;
set sample.a;
where bidprice or askprice or QUALIFIERS is not missing;
run;

/* sample.c*/
data sample.c;
set sample.b;
second=second(time);
minute= minute(TIME);
hour=hour(time);
day=day(date);
month=month(date);
year=year(date);
run;


/*  SAMPLE.E,*/
data sample.d;
set sample.c;
sum+1;
output;
run;


/**/
proc sort data= sample.d;
by ric month day;
run;

data sample.e;
set sample.d;
by date;
if first.date=1 then do;
num+1;
output sample.e ;
end;
else do
num=num;
output sample.e;
end;
run;


data sample.sort1_&name&number;
set sample.e;
run;

data sample.e;
set sample.e;
 if QUALIFIERS='O[GV1_TEXT]' then delete;
 if QUALIFIERS='f[GV1_TEXT]' then delete;
 if QUALIFIERS='p[GV1_TEXT]' then delete;
 if QUALIFIERS='e[GV1_TEXT]' then delete;
 run;
```

```
/*, ,sample*/

data sample.trades;
set sample.trades;
delete;
run;

%sample;

proc sort data = sample.trades;
by date;
quit;

data sample.trades;
set sample.trades;
by date;
if first.date=1;
run;

data sample.trades;
set sample.trades;
where  buys and sells is not missing;
run;


/**/
proc append base = sample.trades_sum data = sample.trades;
run;



/*pin*/

ods output AdditionalEstimates=sample.pin
ConvergenceStatus=sample.pincs
 IterHistory=sample.pinih FitStatistics=sample.pinfs;

proc nlmixed data=sample.trades fd=central technique=quanew
update=bfgs;
 by ticker;
 parms a=.1 .5 .9, d=.1 .5 .9, u=20 200 2000, e=20 200 2000;
 bounds 0 <= a d <= 1, u e >= 0;

 pin = a*u / (a*u + 2*e);

 temp = (1-a)*pdf('poisson',buys,e)*pdf('poisson',sells,e)
        + a*d*pdf('poisson',buys,e)*pdf('poisson',sells,u+e)
        + a*(1-d)*pdf('poisson',buys,u+e)*pdf('poisson',sells,e);

 if temp = 0 then temp = 1E-300;

 loglik = log(temp);

 model buys~general(loglik);

 estimate 'alpha' a;
 estimate 'delta' d;
 estimate 'mu' u;
 estimate 'epsilon' e;
```

```
 estimate 'PIN' pin;
run;

proc print data=sample.pin label;
label ticker='Stock ticker';
title 'PIN estimates';
run;

proc print data=sample.pincs;
title 'Convergence Status for MLE procedure';
run;

proc print data=sample.pinfs;
title 'Additional statistics';
run;


/*pin */
proc append base = sample.pin_sum data = sample.pin;
run;


/*PIN*/
data sample.pin_sum2 ( keep = ticker label estimate);
set sample.pin;
run;

proc transpose data=sample.pin_sum2 out = sample.pin_sum3;
id label;
by ticker;
run;

data sample.pin_sum3;
set sample.pin_sum3;
drop _NAME_;
run;

proc append base= sample.pin_sum_end data =sample.pin_sum3;
run;




/*adjpin*/
ods output AdditionalEstimates=sample.adjpin
ConvergenceStatus=sample.adjcs
 IterHistory=sample.adjih FitStatistics=sample.adjfs;


proc nlmixed data=sample.trades fd=central technique=quanew
update=bfgs;
 by ticker;
 parms a=.1 .5 .9, d=.1 .5 .9, t=.1 .5 .9,us=20 200, ub=20 200, es=20
200, eb=20 200, dels=20 200 , delb=20 200 ;
 bounds 0 <= a d t <= 1, us ub es eb  dels delb >= 0;

adjpin = a*(d*ub+(1-d)*us) / (a*(d*ub+(1-d)*us)+ (delb+dels)*(a*t+(1-
a)*t) + es + eb);

psos = (delb+dels)*(a*t+(1-a)*t) / (a*(d*ub+(1-d)*us)+
(delb+dels)*(a*t+(1-a)*t) + es + eb);
```

```
temp =  (1-a)*(1-t)*pdf('poisson',buys,eb)*pdf('poisson',sells,es)
+ (1-a)*t*pdf('poisson',buys,eb+delb)*pdf('poisson',sells,es+dels)
+a*(1-t)*(1-d)*pdf('poission',buys,eb)*pdf('poisson',sells,us+es)
+a*t*(1-d)*pdf('poission',buys,eb+delb)*pdf('poisson',sells,us+es+dels)
+a*(1-t)*d*pdf('poission',buys,ub+eb)*pdf('poisson',sells,es)
+a*t*d*pdf('poission',buys,ub+eb+delb)*pdf('poisson',sells,es+dels);

 if temp = 0 then temp = 1E-300;
 loglik = log(temp);

 model buys~general(loglik);

 estimate 'alpha' a;
 estimate 'delta' d;
 estimate 'theta' t;
 estimate 'mu_sell' us;
 estimate 'mu_buy' ub;
 estimate 'epsilon_buy' eb;
 estimate 'epsilon_sell' es;
 estimate 'delta_buy' delb;
 estimate'delta_sell' dels;
 estimate 'PSOS' psos;
 estimate 'adjPIN' adjpin;

run;

proc print data=sample.adjpin label;
 label ticker='Stock ticker';
 title 'adjPIN estimates';
run;

proc print data=sample.adjcs;
 title 'Convergence Status for MLE procedure';
run;

proc print data=sample.adjfs;
 title 'Additional statistics';
run;

proc append base = sample.adjpin_sum data = sample.adjpin;
run;

data sample.adjpin_sum2 ( keep = ticker label estimate);
set sample.adjpin;
run;

proc transpose data=sample.adjpin_sum2 out = sample.adjpin_sum3;
id label;
by ticker;
run;

data sample.adjpin_sum3;
set sample.adjpin_sum3;
drop _NAME_;
run;

proc append base= sample.adjpin_sum_end data =sample.adjpin_sum3;
run;

   %end;
```

```
%mend company;

%company(90, companyname);


proc sort data= sample.pin_sum_end out= pin_sort_temp;
by ticker;
quit;

data sample.pin_sort;
set pin_sort_temp;
by ticker;
if first.ticker=1;
run;

proc sort data= sample.adjpin_sum_end out= adjpin_sort_temp;
by ticker;
quit;

data sample.adjpin_sort;
set adjpin_sort_temp;
by ticker;
if first.ticker=1;
run;


data sample.pin_adjpin_qualifier;
merge sample.adjpin_sort sample.pin_sort;
keep adjpin psos pin ticker;
run;



/*option

data sample.trades_justify_remove sample.trades_justify;
set sample.trades_sum;
sellbuysum = sells+buys ;
if sellbuysum < 30  then do;

output sample.trades_justify_remove;
end;
else output sample.trades_justify;
run;

proc sort data=sample.trades_justify;
by ticker;
quit;

proc means data= sample.trades_justify noprint MEAN ;
output out = sample.trades_justify_mean;
var sellbuysum;
by ticker;
run;


data sample.trades_justify_result(drop=_TYPE_ _STAT_ _freq_
sellbuysum);
set sample.trades_justify_mean;
where _STAT_ in('MEAN');
if _freq_ >60;
```

```
run;



proc sort data=sample.Pin_adjpin_qualifier;
by ticker;
quit;

proc sort data=sample.trades_justify_result;
by ticker;
quit;


data sample.Pin_adjpin_qualifier_remove;
merge sample.Pin_adjpin_qualifier(in = s)
sample.trades_justify_result(in=t);
by ticker;
if t=1;
run;

*/

Run;
```

```
/*programme 4, combining PIN and Adjpin with firm trading and
accounting data , statistical summary, simple correlation, OLS
regression*/

libname sample 'C:\Documents and Settings\LabUser\Desktop\cleaned firm
characteristic data';
run;

proc printto print="C:\Documents and Settings\LabUser\Desktop\cleaned
firm characteristic data\output.txt";
run;

proc printto log ="C:\Documents and Settings\LabUser\Desktop\cleaned
firm characteristic data\log.txt";
run;

*proc printto;
*run;
/*
proc sort data=sample.Pin_adjpin_nonqualifier_remove;
by ticker;
quit;

data sample.regdata_nonqualifier;
merge sample.Pin_adjpin_nonqualifier_remove(in = s)
sample.companyyearlydata(in=t);
by ticker;
if t=1;
run;

proc corr data= sample.regdata_nonqualifier
out=sample.regdata_nonqualifier_correlation pearson spearman;
var pin adjpin psos ROA STDEV LOGMV amihud turnover TOBINQ;
quit;


*/


%macro univariate (file, variable);
proc univariate data= sample.&file;
var &variable ;
HISTOGRAM &variable /NORMAL;
PROBPLOT &variable ;
output out= sample.uni_&file._&variable N = n SKEWNESS=skewnesss
KURTOSIS=kurtosis  MEAN=mean STD=std  MAX=max MIN=min Q1=Q1 MEDIAN=
median Q3=Q3;
title ' unvariate test:  N SKEWNESS KURTOSIS  MEAN STD CLM  MAX P25
P50 P75 MIN (qualifer)';
run;

%mend;



data sample.Pin_adjpin_qualifier_remove;
attrib ticker length= $40 ;
attrib PSOS  informat = best12. format=best12.;
attrib adjPIN informat = best12. format=best12.;
attrib PIN informat = best12. format=best12.;
set sample.Pin_adjpin_qualifier_remove;
```

```
run;


proc sort data=sample.Pin_adjpin_qualifier_remove;
by ticker;
quit;

proc sort data= sample.companyyearlydata;
by ticker;
quit;


data sample.regdata_qualifier;
merge sample.Pin_adjpin_qualifier_remove(in = s)
sample.companyyearlydata(in=t);
by ticker;
if t=1;
run;


proc corr data= sample.regdata_qualifier
out=sample.regdata_qualifier_correlation pearson spearman;
var pin adjpin psos ROA STDEV LOGMV amihud turnover TOBINQ logprice;
quit;


proc means data= sample.regdata_qualifier N SKEW KURT  MEAN STD CLM
MAX P25 P50 P75 MIN;
output out=sample.regdata_qualifier_mean;
var pin adjpin psos ROA STDEV LOGMV amihud turnover TOBINQ logprice;
title '  N SKEW KURT  MEAN STD CLM  MAX P25 P50 P75 MIN (qualifer) ';
run;


%univariate (regdata_qualifier, pin);
%univariate (regdata_qualifier, adjpin);
%univariate (regdata_qualifier, psos);
%univariate (regdata_qualifier, ROA );
%univariate (regdata_qualifier, STDEV );
%univariate (regdata_qualifier, LOGMV );
%univariate (regdata_qualifier, amihud );
%univariate (regdata_qualifier, turnover );
%univariate (regdata_qualifier,TOBINQ );
%univariate (regdata_qualifier,logprice);



proc reg data=sample.regdata_qualifier
outest=sample.reg_q_adjpin_roa_std_mv_ami_tob OUTSEB COVOUT TABLEOUT ;
model adjpin =   ROA STDEV LOGMV amihud TOBINQ/ edf ;
run;
quit;

proc reg data=sample.regdata_qualifier
outest=sample.reg_q_pin_roa_std_mv_ami_tob OUTSEB COVOUT TABLEOUT;
model pin =   ROA STDEV LOGMV amihud  TOBINQ/ edf ;
run;
quit;
```