125899

MBS (Finance) Thesis

Cross-sectional analysis of pricing efficiency, liquidity, and information asymmetry

Student: Xin Gao

ID: 04121198

Supervisor: Dr Nuttawat Visaltanachoti

Date: 7 December, 2009

Table of Contents

Acknowledgements

I am thankful to my supervisor, Dr Nuttawat Visaltanachoti, for great encouragement, with constructive suggestions and comments. I could not have completed this study without his great help. I would also like to thank Associate Professor John Pinfold, Professor Ben Jacobsen, Dr Liping Zou, and Dr Mei Qiu for their valuable suggestions. All errors are my own responsibility.

Contact information:

Tel: +64-9-4190715

E-mail address: x.g.gao@hotmail.com

Abstract

This paper tests the relation between pricing efficiency and liquidity, with and without, the effects of asymmetric information. First, we show that informed trading is negatively related to liquidity. This result is consistent with previous researches, which find that informed trading reduces liquidity. Second, this report explores the direct relation between price efficiency and liquidity by applying a cross-sectional regression. The result indicates that liquidity associated with asymmetric information effects enhances pricing efficiency. The cross-sectional relation between relative informational efficiency and liquidity combed with informed trading is significantly positive. Third, we find that pure liquidity trading also contributes to price informativeness. The positive relation between relative informational efficiency and liquidity unrelated to asymmetric information cannot be rejected.

1. Introduction

This study examines the relation between pricing efficiency, liquidity, and information asymmetry. It has been argued in the literature that liquidity is an important factor in determining the degree of price informational efficiency. Previous researches focus on the indirect relation between pricing efficiency and liquidity. Liquidity indirectly affects pricing efficiency through the activity of informed traders, who make prices more efficient. High liquidity and low trading friction reduces the trading costs of informed traders and encourages them to trade. Liquidity is the capability to purchase, or sell, a large quantity of securities quickly, with a comparatively small price impact (Campbell, Lo & Mackinlay, 1997). Pricing efficiency is defined by two stipulations: First, the stock return follows a random walk; and second, a stock price must fully reflect all available information in the market. Regulators and market participants believe that an increase in liquidity has a positive impact on pricing efficiency.

Informed trading activities are positively related to the extent of concentration of liquidity trading: The more concentrated the liquidity trading, the more active the informed trading (Admati & Pfleiderer, 1988). The participation of informed traders makes prices more informative. Harris (2003) explains that informed traders estimate a stock's fundamental value based on private information that only they have access to, or on public information that everyone can acquire. They purchase stocks that they think are considerably undervalued and sell stocks that they think are considerably overvalued, in order to make profits. As they purchase stocks that are priced below their estimated fundamental values, and short them otherwise, their trades converge the stock prices to their estimated intrinsic value. Their trading costs are the price impacts of their participation. In order to make more profit, they are likely to trade in a liquid market, where prices are significantly different from the true value of the stocks. Leland (1992) finds that, when insider trading [informed trading based on material information that traders obtain directly, or indirectly, from the management of a firm, and that is not yet announced publicly (Harris, 2003)] is allowed, stock prices are more informative and insiders will profit. Aktas, de Bodt and van Oppens (2008) study the sensitivity of daily returns to the relative order imbalance (the ratio of the difference in the of number of daily buys and sells, to the total number of buys and sells) made by insider trading and find that insiders dramatically stimulate price discovery on insider trading days. Keiber

(2006) argues that, if there is no insider trading activity, a stock price deviates from its fundamental value and the variation in the price is large. The empirical results also show that an improvement in liquidity indirectly boosts pricing efficiency, because high liquidity and low trading frictions decline the trading cost of informed traders, and stimulate them to trade (see Admati & Pfleiderer, 1988; Chordia, Roll & Subrahmanyam, 2008; Roll, Schwartz & Subrahmanyam, 2007). Subrahmanyam (2009, p. 527) also asserts that "...both liquidity and liquidity risk are priced in the cross-section of stock returns, the law of one price is more likely to hold in more liquidity markets, and liquidity enhances market efficiency. Hence policies to enhance liquidity encourage efficiency and reduce costs of raising capital".

The counter-view states, however, that because informed trading induces adverse selection cost, it reduces market liquidity and harms the market. Since informed traders have an informational advantage over the market-maker in regards to a stock's fundamental value, trading with such investors will mean potential losses for the market-maker and other uninformed traders. Furthermore, because market-makers and uninformed traders are not able to differentiate the informed from the uninformed, they need compensation for the risk of trading with such traders, through enlarging the bid-ask spread (Campbell et al., 1997; Kyle, 1985). Moreover, since an increase in liquidity trading also induces more informed traders (see Admati & Pfleiderer, 1988; Kyle, 1984, 1985; Subrahmanyam, 1991) and under the assumption that informed traders are risk averse, their cumulative expected profit may increase in response to the entry of more informed traders, meaning that increasing liquidity reduces the market liquidity (Subrahmanyam, 1991). Fishman and Hagerty (1992) also argue that informed trading can harm the market in two ways. First, the small number of better informed traders prevents non-informed traders from obtaining information and trading. Second, the uneven distribution of information among the traders; i.e., asymmetric information effects; results in a less competitive market.

Other opinions argue that if the market makers are performing well, such as balancing the bid-ask spreads at an appropriate level and there being no misquoted stock (Chordia et al., 2008), and/or the fundamental value is well-known, such that the price will be informative without informed trading (Harris, 2003), then liquidity is not relevant to pricing efficiency, or increasing liquidity, leaves the market efficiency unchanged (Kyle, 1984, 1985).

This study intends to solve this puzzle through empirical evidence. The first hypothesis is that there is a significantly positive relation between price efficiency and liquidity combined with informed trading in the cross-sectional regression in the intraday pattern. Recently, similar testing has been carried out by Chordia, Roll and Subrahmanyam (2008), who use time series to indicate the indirect relation between liquidity and return predictability. They observe that illiquidity decreases over time, based on five-minute intervals covering the period between January 1993 and July 2002. Meanwhile, return predictability of the lagged (by five minutes) dollar order imbalance declined in the same sample period. The coefficients of the daily return-variance ratio and the autocorrelation approaches a random walk when there is an improvement in liquidity and a reduction in the minimum tick size. Unlike Chordia, Roll and Subrahmanyam (2008), however, we contribute a method to explore the direct relation between price efficiency and liquidity by applying a cross-sectional regression. First, we follow Boehmer, Kelley and Hall (2009) to calculate return variance ratios, autocorrelations, and pricing errors, as proxies for relative informational efficiency, for each company, at five-, ten-, and thirty-minute intervals. We assume that, if prices are efficient, they should follow a random walk. The relative informational (price informativeness) efficiency is, then, defined as how closely transaction prices resemble the efficient prices, i.e. a random walk. Second, we follow Chordia, Roll and Subrahmanyam (2000) and Pukthuanthong-Le and Visaltanachoti's (2009) approach to compute the average effective spreads, absolute spreads, and relative spreads, as measures of liquidity for each firm in the intraday pattern. Then, we investigate whether liquidity explains the cross-sectional variation in efficient prices. The significant relation between relative pricing efficiency and liquidity shows support for our first argument, that liquidity combined with informed trading has positive impact on price informational efficiency.

Additionally, market regulators improve liquidity through enforcements of insider trading laws to reduce the adverse selection costs, and reductions in the minimum tick size to decrease trading costs. Prohibiting insider trading is likely to improve the liquidity, at the cost of informed traders. Moreover, on January 29, 2001, the NYSE market executed decimalisation (one cent increment/decrement) on stock trading orders. Empirical results show that transaction costs reduced significantly since decimalisation, and liquidity had been improved substantially (see Bessembinder, 2003b; Chakravarty, Wood & van Ness, 2004; Furfine, 2003). It may, however, increase front running problems. Front runners acquire information about orders that other investors have chosen to submit. They then attempt to trade on their own account before those investors settle their trades. Since the orders of front runners are submitted first, they profit instantly from price movements of the following orders of other investors. If the investors they front run are liquidity traders who trade for exogenous reasons, front runners usually cause prices to be less informationally efficient, since liquidity traders usually participate in the market for other reasons rather than information about stock intrinsic values (Harris, 2003). Decimalisation encourages front running activities, since front runners can trade before other investors at much cheaper cost. Therefore, the second aim of this research is to test whether pure liquidity trading has direct impacts on stock pricing efficiency. Although Boehmer, Kelley and Hall (2009) find that institutional holding and trading activities in the absence of private information can enhance market informational efficiency, if traders are able to differentiate between permanent changes (due to the revelation of new information) in stock prices and temporary deviations (pricing inefficiency due to the trades for exogenous reasons) from a fundamental stock value, to the best of our knowledge there is no other paper studying the direct relation between stock pricing efficiency and pure liquidity trading that is unrelated to private information. Exploring this relation can help market regulators obtain insight as to whether improving pure liquidity can actually contribute to price informativeness.

One approach of measuring liquidity without asymmetric information effects is based on the model developed by Glosten and Harris (1988). They use a linear regression to estimate a model of the bid/ask spread. There are two components in this liquidity model: The adverse selection component (asymmetric information); and the transitory component, which includes inventory costs, clearing fees, and/or monopoly profits. Since we are only interested in decomposing information asymmetry from liquidity, we simplify the model into a linear regression model by including only two variables: Liquidity; and information asymmetry. The three spread measures mentioned above are used as proxies for liquidities. We follow Duarte and Young (2009) in using adjusted-PIN as a measure of pure information asymmetry. In Duarte and Young's (2009) model, they separately decompose the probability of informed trading (PIN) into two components: Information asymmetry; and illiquidity. They then adopt adjusted-PIN (adjPIN) as a proxy for information asymmetry, and the probability of symmetric order-flow shock (PSOS) as a measure of illiquidity that is not related to information asymmetry. In our simplified liquidity model, we use residual values, or error series {e}, to capture all other effects on liquidity that are not driven by information asymmetry. Another measure for pure liquidity is the probability of symmetric order-flow shock, which can be observed directly from Duarte and Young's (2009) model.

Note that this paper does not actually test pricing efficiency. Rather, this research focuses on the impact of liquidity with, and without, information asymmetry on the degree of pricing efficiency. Therefore, such measures for relative pricing efficiency are not necessary to capture the pricing error in an absolute sense.

The rest of this paper is organised as follows. Section 2 outlines the background of research into liquidity, information asymmetry, and their impacts on stock pricing efficiency. Section 3 describes the data used in our research. Sections 4 and 5 outline the methods applied and the statistical results of the measures of pricing efficiency and liquidity with, and without, asymmetric information effects. Section 6 shows the empirical results regarding the cross-sectional relation between pricing efficiency and liquidity related, and unrelated to, private information. Sections 7 and 8 discuss the limitations and conclusions of the research, respectively.

2. Literature Review

Liquidity is the capability to purchase, or sell, a large amount of securities quickly, with comparatively little price impact (Campbell et al., 1997). Measures of liquidity and its impacts on a security's price have been well studied in the literature. Common measures include bid-ask spreads, trade size, firm size, and asymmetric information risk. Among these measures, information asymmetry is a determining factor of liquidity and a main cause of price impacts in the findings of recent studies.

The bid-ask spread measure has two broad components: The adverse selection component (asymmetric information); and the transitory component, which includes inventory costs, order-processing costs/clearing fees, and/or monopoly profits (Campbell et al., 1997; Glosten & Harris, 1988). An increase in both adverse selection costs and transitory costs result in a positive bid-ask spread (Amihud & Mendelson, 1980; Glosten & Milgrom, 1985). Since informed traders have informational advantage about a stock's fundamental value [fundamental value is the true value of the stock (Harris, 2003)] that the market maker does not, trading with such investors will lead to potential losses for the market maker. Moreover, because market makers are not able to differentiate the informed from the uninformed, they need compensation for the risk of trading with such traders by enlarging the bid-ask spread (Campbell et al., 1997; Kyle, 1985).The bid-ask spread is positively related to stock returns (Amihud & Mendelson, 1986).

The trade size also reveals the impact of liquidity on price. Trade size has a significantly positive impact on price movement, due to the behaviour of informed traders. Informed traders like to trade a large quantity at any given price (Easley & O'Hara, 1987; Keim & Madhavan, 1996). The impact of trades on spread size is also positive, as large trades result in a large bid-ask spread. A larger price impact is found in trading those stocks with relatively wide spreads, in comparison with those that have small spreads (Hasbrouck, 1991a).

Firm size is also a proxy for liquidity. A large firm's stock is traded more frequently and has a relatively narrower bid-ask spread and less price impact/return than does a small firm (Amihud & Mendelson, 1986; Banz, 1981; Fama & French, 1992). Small firms are associated with more considerable price impacts and degrees of information asymmetry (Hasbrouck, 1991a).

Asymmetric information risk is an important, and increasing, function of illiquidity as it induces adverse selection problems. Easley, Kiefer, O'Hara and Paperman (1996) test the impact of information-based trading on the differences in spreads for active and inactive traded stocks by using a model referred to as the Probability of Informed Trading model (PI or PIN model). They use intraday trading data to investigate the frequency of new information events, the composition of trading when new information arrives in the market, and market depth for securities with diverse volume-deciles. Their main results indicate that actively traded stocks are associated with a low probability of informed trading. Larger spreads and returns for infrequently traded stocks arise from greater informed trading risks, rather than being the result of market power, or an inventory effect. They imply that private information is more crucial for inactively traded stocks. Arrival of new information brings larger price effects to inactive stocks, than to active ones.

Chen, Chung, Lee and Liao (2007) provide another evidence to support information asymmetry being positively related to illiquidity, by analysing the level of corporate governance. They find that liquidity providers tend to increase the spread to those firms with higher levels of information asymmetry risk, and that such price-protection actions will decrease the stock market liquidity. In contrast, companies with better corporate governance and disclosure practices have higher market liquidity (lower bid-ask spreads), because of the decrease in information asymmetry risk in those firms.

Affleck-Graves, Callahan and Chipalkatti (2002) explore the relation between earnings predictability, information asymmetry, and market liquidity. Their results indicate that, first, on the trading of and the trading day prior to, the announcements of quarterly earnings, there is an increase in the adverse selection cost of the bid-ask spreads for those firms with less earning predictability. Such phenomenon does not appear in firms with highly predictable earnings. Second, firms with less predictable earnings have constantly high bid-ask spreads compared with those with high predictable earnings across time. As a result, they conclude that stocks with less predictable earnings cause an increase in the information asymmetry between privately informed investors and dealers (henceforth, market-makers) in the stock market. Market-makers, again, need compensation for this information asymmetry by increasing the bid-ask spreads for those firms with less predictable earnings.

In contrast, from an informed trader point of view, high liquidity and low trading friction reduce the trading costs of informed traders and encourage them to trade. Admati and Pfleiderer (1988) study the behaviour of liquidity traders and informed traders in the financial market, and find that both liquidity trading (directly) and informed trading (indirectly) contribute to the total trading volume. This is because the intense of liquidity trading is an influential factor for both uninformed traders and informed traders to participate in the market (also see Kyle, 1984, 1985; Subrahmanyam, 1991).

There are also other measures and implications of liquidity. Amihud (2002) measures illiquidity as the ratio of the absolute value of the average daily return of a stock, to its dollar volume. He finds that the market expected illiquidity has a positive impact on ex ante stock excess returns. In contrast, there is a negative relation between contemporary unexpected illiquidity and stock returns. Chordia, Roll and Subrahmanyam (2002) compute the accumulated daily order imbalance as the buy initiated, less than sell initiated, in terms of order, share, and dollar values. They find that order imbalances in either direction have a negative impact on liquidity. There is also a significant relation between current and lagged order imbalance and market returns, in that "…signed order imbalances are high following market declines and low following market advances" (Chordia et al., 2002, p. 128). Chordia, Roll and Subrahmanyam (2000) and Pukthuanthong-Le and Visaltanachoti (2009) adopt bid-ask spread, proportional spread, depth, effective spread, imbalance of depth, and slope of outside quotes, to explore the commonality in liquidity in the NYSE and the Stock Exchange of Thailand (SET), respectively. They find that there is co-movement between individual stock liquidity and market- and industry-wide liquidity. Moreover, Pukthuanthong-Le and Visaltanachoti (2009) find that the co-movement between individual liquidity and industry-wide liquidity is stronger than the co-movement with market-wide liquidity in the SET.

The effects of liquidity combined with informed trading on stock pricing efficiency is non-monotonic. Pricing efficiency is defined in that: First, the stock return follows a random walk; and second, a stock price must fully reflect all available information.

There are several common measures used to investigate pricing efficiency: Testing of the variance ratio and use of the return autocorrelation to exam the random walk hypothesis - there is no serial dependence of stock returns for all lags and leads (Boehmer et al., 2009; Campbell et al., 1997; Lo & MacKinlay, 1988); using the modem method, which consists of both return and information related factors, to capture the deviation from the efficient price (Boehmer et al., 2009; Hasbrouck, 1991a, 1991b, 1993); and/or analysing the predication power (R-squared) of the explanatory variable, order imbalance, on returns in the intraday pattern (Chordia, Roll & Subrahmanyam, 2005; Chordia et al., 2008).

On the one hand, participation of informed traders makes price more informative. Informed traders estimate a stock fundamental value based on private information that only they have access to, or based on public information that everyone can acquire. They purchase a stock that they think is considerably undervalued and sell a stock that they think is considerably overvalued, in order to make a profit. As they purchase a stock which is priced below their estimated fundamental value and short it otherwise, their trades converge the stock price to the estimated intrinsic value. Their trading costs are the price impacts of their participation. In order to make more profit, they are likely to trade in a liquid market, where prices are significantly different from the true value (Harris, 2003). Leland (1992) finds that when insider trading [informed trading based on material information that traders obtain directly, or indirectly, from the management of a firm and that is not yet announced publicly (Harris, 2003)] is allowed, stock prices are more informative and insiders will profit.. Aktas, de Bodt and van Oppens (2008) examine the sensitivity of daily returns to the relative order imbalance (the ratio of the difference in the number of daily buys and sells, to the total number of buys and sells) made by insiders trading, and find that insiders stimulate price discovery dramatically on insider trading days. Keiber (2006) argues that if there are no insider trading activities, a stock price deviates from its fundamental value and the variation in price is large. An improvement in liquidity boosts pricing efficiency, because high liquidity and low trading frictions decline trading cost of informed traders, encouraging them to trade (see Admati & Pfleiderer, 1988; Boehmer et al., 2009; Chordia et al., 2008; Roll et al., 2007). Subrahmanyam (2009, p. 527) also asserts that "...both liquidity and liquidity risk are priced in the cross-section of stock returns, the law of one price is more likely to hold in more liquidity markets, and liquidity enhances market efficiency. Hence policies to enhance liquidity encourage efficiency and reduce costs of raising capital".

On the other hand, asymmetric information induces adverse section costs, which is negatively related to market liquidity. Allowing insider trading reduces the total market liquidity, and outside investors and liquidity (uninformed) traders will face losses (Leland, 1992). Fishman and Hagerty (1992) argue that insider trading can harm the market in two ways. First, the small number of better informed traders prevents non-informed traders from obtaining information and trading. Second, the uneven distribution of the information among the traders, i.e., asymmetric information effects, results in a less competitive market. Moreover, because an increase in liquidity trading also induces more informed traders (see Admati & Pfleiderer, 1988; Kyle, 1984, 1985; Subrahmanyam, 1991), increasing liquidity leaves the market efficiency unchanged (Kyle, 1984, 1985), or under the assumption that informed traders are risk averse, their cumulative expected profit may increase in response to the increased entry of informed traders, meaning that increasing liquidity reduces the market liquidity (Subrahmanyam, 1991). Increasing liquidity also increases front running problems. Front runners acquire information about orders that other investors have chosen to submit. They then attempt to trade on their own account, in front of those investors. Since the order of front runners is submitted first, they profit instantly from the price movement of the following orders of the other investors that they front run. If the investors they front run are liquidity traders, the front runners usually make the price less informationally efficient, since liquidity traders' participation in the market is not usually based on information about stock fundamental values. If the investors they front run are informed, then the front runners make the price more informative in the short-term. These trades converge the price to the intrinsic value more quickly in the short-run. Front running practices usually reduce the market liquidity, because profits made by front runners are transaction costs for the other traders that they front run. Investors that face front running problems are forced to take a worse price than where there are no front running activities (Harris, 2003).

More recently, Duarte and Young (2009) test whether a proxy for the probability of informed trading (PIN) is priced due to the influence of asymmetric information, or due to other liquidity effects that are not related to asymmetry information, using NYSE listed stocks from 1983 to 2005. Their model is based on that of Easley et al.'s (1996)

PIN model. Additionally, Duarte and Young (2009) allow large variation of the buy/sell order flows and simultaneous order flow shocks in their extended model. They decompose PIN into two separate components: Information asymmetry; and illiquidity. They use adjusted PIN (adjPIN) as a proxy for information asymmetry, and the probability of symmetric order-flow shock (PSOS) as a measure of illiquidity that is not related to information asymmetry. Their findings suggest that high PSOS (low liquid) firms tend to have higher PINs than other firms, and that firms with high PSOS have only slightly higher information asymmetry than firms with low PSOS. Moreover, PSOS (illiquidity) has a strong and positive relation with expected returns. In contrast, information asymmetry does not have a statistically significant effect on expected returns. Therefore, they conclude that PIN is priced because of liquidity effects that are unrelated to the information asymmetry. Such effects explain the relation between PIN and the cross-sectional expected returns. Information asymmetry, however, does not have explanatory power on either PIN, or expected returns (Duarte and Young, 2009).

3. Data

All NYSE listed common stocks during the period between 3 January, 2007 and 28 March, 2007 are obtained from the Center for Research in Security Prices (CRSP) and the Reuters databases. Since stocks are priced efficiently over a daily time interval, this research focuses on pricing efficiency in the intraday pattern, with the data including intraday trading quotes, the size of bid (ask) orders, transaction prices, and trading volume. Sixty trading days' data is adequate to estimate the parameters of PIN and pricing errors (Easley et al., 1996; Hasbrouck, 1991a, 1991b, 1993). Research based on 5-minute intervals will provide the most appropriate results for two reasons: First, non-trading activities would impose an problem in shorter intervals; and second, market inefficiencies would be less obvious in longer intervals (Chordia et al., 2008). This paper reasonably assumes that there is no reporting delay since 1998 (for example, see Bessembinder, 2003a; Boehmer et al., 2009; Chordia et al., 2008). As a result, the first quote will be simply defined as the one prior to the trades. Also, there are several other important criteria. The selected firms must be listed in the NYSE during the years from 1996 to 2008 in order to avoid potential IPO and bankruptcy impacts. Then, those firms that moved from the NYSE to another market are excluded in order to avoid impacts resulting from variations in the market microstructure (Amihud, 2002; Reinganum, 1990). Firms that did not trade at least once a week during the sample period are also excluded. These criteria narrow the firms down to 1,044 in total. Only quotes and transactions happening within trading hours will be included. To avoid potential errors, this research purges the data by deleting data for which one or more of the following occurs (Sarkar & Schwartz, 2009):

- Trade price is zero, negative, or missing;
- Quote of buy (sell) or size of buy (sell) order is zero, negative, or missing; or
- Volume is zero, or missing.

This paper applies Lee and Ready's (1991) approach to identify buyer initiated and seller initiated trades. Buyer initiated will be defined as those trades above the bid and ask quote mid-point; while seller initiated are those trades that are below the bid and ask quote mid-point. If a trade happens at the quote mid-point, we apply the tick test to identify whether it is buyer initiated, or seller initiated [that is, when comparing the trade with the previous price, if the current price is above (below) it, a trade is considered to be buyer-initiated (seller-initiated)]. We also sum the number of buyer initiated and seller initiated trades on each given day.

Table 1 shows the median and the percentiles of the cross-sectional series of daily number of buys and sells initiated for each firm, based on the 1,044 firms during the period of January 3, 2007 to March 28, 2007. This table implies that there is a greater number of daily buys, than daily sells in all percentiles. Also, a large variance in buyer initiated and seller initiated is found, with medians of 87,047 and 68,454, respectively. The high volatility of buys and sells is due to the different believes on information among investors (Sarkar & Schwartz, 2009). Such a finding is also consistent with Duarte and Young's (2009) result that buys are more volatile than sells, and that there is a large variance in both the buys and sells initiated.

Table 1 also presents the median and the percentiles of the cross-sectional series of Pearson correlations between the average daily numbers of buys and sells for each firm. We find that there is a significantly positive correlation between buys and sells, with the median coefficient of 0.79 (p-value <0.0001). This result implies that there is a symmetric order flow shock in the market trading. An increase (decrease) in the number of buys is associated with a simultaneous increase (decrease) in the number of sells. Sarkar and Schwartz (2009) explain that investors participate in the market based on information they observe. Different interpretations on public information among traders (Harris & Raviv, 1993; Kandel & Pearson, 1995), or diverse information signals that investors observe (He & Wang, 1995) are a trigger of symmetric order flows.

Overall, we find that a large variance in the daily number of buys and sells exists, and that the average daily number of buys is positively related to the average daily number of sells.

Table 1: **Summary Statistics on the Number of Buys and Sells Initiated**

This table shows the median and the percentiles of the cross-sectional distribution of a series of statistics on the average daily number of buys and sells initiated for each firm based on 1,044 firms during the period of January 3, 2007 to March 28, 2007. Intraday data is obtained from the CRSP and Reuters databases. Buy- and sell- initiated are identified by the approach of Lee and Ready (1991).

Variable	5th Pctl	25th Pctl	Median	95th Pctl	95th Pctl
Mean buys	90	393	730	1329	3241
Mean sells	85	352	639	1200	3045
Variance buys	2177	28008	87047	238990	967687
Variance sells	1908	22385	68454	198094	852511
Correlation between buys and sells	0.42	0.67	0.79	0.87	0.94
P-value	<0.0001	<0.0001	<0.0001	<0.0001	0.0010

4. Pricing Efficiency

This paper adopts the approach used in Boehmer, Kelley and Hall (2009), using the following as proxies for relative pricing efficiency; variance ratios, return autocorrelations, and pricing errors. The relative informational (price informativeness) efficiency is defined as how closely transaction prices resemble the efficient prices, i.e. a random walk. Note that all measures adopted in this section are inversed measures of relative pricing efficiency.

The variance ratios, VR(n, m), are calculated as the ratio of the mid-quote return variance over m periods against the return variance over n period (Lo & MacKinlay, 1988). Assume that prices follow a random walk in an absolutely efficient market. With an variance ratio equal to 1, the inverse indicator of market efficiency adopted in this paper is /1 – VR(n, m)/, which is the absolute value of the gap between the actual and efficient prices (Boehmer et al., 2009).

To calculate the mid-quote return autocorrelations, this paper also assumes that, if the prices are efficient, it should follow the random walk (Boehmer et al., 2009), and that the coefficients of the first order autocorrelation should be zero at all leads and lags (Campbell et al., 1997). The coefficient of the first order mid-quote return autocorrelation in the n-minute interval is the autocovariance of the mid-quote return and its lagged mid-quote return divided by the variance of returns. The absolute value of the 10-minute and 30-minute mid-quote returns' autocorrelation coefficients, /AR 10/ and /AR 30/, respectively, are used as inverse indicators of relative pricing efficiency.

The model used here for calculating pricing errors is first contributed by Hasbrouck (1993), and is adopted by Boehmer, Kelley and Hall (2009) as an inverse measure of informational efficiency. They define that the (log) transaction price, p_t, equals the random walk component, m_t, plus a transitory pricing error, s_t, where t is the transaction time:

$$p_t = m_t + s_t \qquad (1)$$

According to Boehmer, Kelley and Hall (2009) and Hasbrouck (1993), the unobservable random walk component, m_t, is also the efficient price. It represents a stock expected value:

$$m_t = m_{t-1} + w_t \tag{2}$$

where m is the random walk component, w are the uncorrelated increments, and t is the transaction time.

The pricing error is the difference between the current price and the efficient price. It may be the result of transaction costs, order imbalances, price discreteness, or dealer inventory effects. Since the pricing error has a zero mean, its standard deviation, σ_s is its magnitude measure. σ_s is the inverse measure of informational efficiency, since it indicates how closely transaction prices resemble the efficient prices, i.e. a random walk. The equation for the variance of pricing error is:

$$\sigma_s^2 = \sum_{j=0}^{\infty} [\alpha_j \quad \beta_j] \, cov(v) \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} \tag{3}$$

We follow Boehmer, Kelley and Hall (2009) and Hasbrouck (1993), in estimating pricing error variance based on a vector autoregression (VAR) consisting of series $\{r_t, x_t\}$, where r_t is the first order trading price difference, and x_t is a vector containing three explanatory trade variables; a trade sign indicator, signed trading volume, and the signed square root of trading volume. $v_{1,t}$ and $v_{2,t}$ are innovations with zero-mean. The VAR equation is:

$$r_t = a_1 r_{t-1} + a_2 r_{t-2} + \cdots + b_1 x_{t-1} + b_2 x_{t-2} + \cdots + v_{1,t}$$

$$x_t = c_1 r_{t-1} + c_2 r_{t-2} + \cdots + d_1 x_{t-1} + d_2 x_{t-2} + \cdots + v_{2,t} \tag{4}$$

The coefficients (α, β) required to compute the pricing error variance are observed from a vector moving average (VMA) representation of VAR, under the BN restriction that s_t must be correlated with the vector series $\{r_t, x_t\}$ (see Beveridge & Nelson, 1981; Hasbrouck, 1993).

$$r_t = a_0^* v_{1,t} + a_1^* v_{1,t-1} + a_2^* v_{1,t-2} + \cdots + b_0^* v_{2,t} + b_1^* v_{2,t-1} + b_2^* v_{2,t-2} + \cdots$$

$$x_t = c_0^* v_{1,t} + c_1^* v_{1,t-1} + c_2^* v_{1,t-2} + \cdots + d_0^* v_{2,t} + d_1^* v_{2,t-1} + d_2^* v_{2,t-2} + \cdots \qquad (5)$$

and

$$\alpha_f = -\sum_{k=j+1}^{\infty} a_k^*$$

$$\beta_f = -\sum_{k=j+1}^{\infty} b_k^* \qquad (6)$$

This paper uses V(s) to refer to σ_s, or the pricing error. The normalised standard deviation, V(s)/V(p), is used as the proxy for relative pricing efficiency, where V(s)/V(p) is the mean pricing error standardised by the standard deviation of the transaction prices, V(p) (Boehmer et al., 2009).

Three points need to be noted here. First, in this paper, five-minute, ten-minute, and thirty-minute intervals are constructed to investigate the serial dependence of a stock quote-midpoint and its lagged value, since converging the market to an efficient level will take more than five, and less than sixty, minutes in the intraday pattern (Chordia et al., 2005). Second, this research sets the first interval quote-midpoint as a missing one when the new day starts. This eliminates the correlation impact of the previous trading day on the new trading day. Third, this paper does not actually test the pricing efficiency. Rather, this research focuses on the impact of liquidity with, and without, information asymmetry on the degree of pricing efficiency. Therefore, measures for relative pricing efficiency are not necessary to capture the pricing error in an absolute sense.

Table 2 shows the cross-sectional means, z-scores, and t-values of the proxies for pricing efficiency. The highest level of pricing error is found in the five-minute interval, with 33.66 (z-value=-20.12) for the thirty- to five-minute variance ratio, and 44.99 (z-value=-14.88) for the sixty- to five-minute variance ratio. Both the coefficients and the significance level of the pricing error decreases as longer time intervals are applied. The coefficients of /1-(VR(10,30)/ and /1-VR(10,60)/ decrease to 16.84 (z-value=-5.27) and 29.53 (z-value=-5.26), respectively. Furthermore, the midpoint return autocorrelation is not significantly different from a random walk in both the ten- and thirty-minute intervals (-0.95 and -0.04, respectively). This finding suggests that pricing error is more obvious in the five-minute interval, and that a stock is priced more efficiently over a longer time interval. Chordia, Roll and Subrahmanyam (2005) explain that it takes more than five minutes, but less than sixty minutes to remove the serial dependence of intraday returns, and to converge stock prices to an efficient level. Market inefficiencies would also be less obvious when time intervals are longer than five minutes (Chordia et al., 2008). The normalised cross-sectional pricing error standard deviation is about 0.39%. This is slightly higher than the NYSE stock average pricing error standard deviation estimated by Hasbrouck (1993), who uses data of 175 firms during the first quarter of 1989, finding a result of 0.33%.

Table 2: Summary Statistics on Proxies of Pricing Efficiency

Table 2 shows the cross-sectional means, z-scores, and t-values of the proxies of pricing efficiency. The computation is based on 1,044 firms listed in the NYSE between January 3, 2007 and March 28, 2007. Intraday data is obtained from the CRSP and Reuters databases. VR(n,m) is the variance ratio of the m period quote midpoint return against the n period quote midpoint return. AR(10) and AR(30) present the autocorrelation of the quote midpoint return in the ten-minute and thirty-minute intervals, respectively. Vs is the pricing error based on Hasbrouck (1993). Vp is the standard deviation of the logged intraday trading prices. The z-scores and t-statistics present the statistical significance (Campbell et al., 1997; Lo & MacKinlay, 1988). Means are expressed as percentages.

	/1-VR(5,30)/	/1-VR(5,60)/	/1-VR(10,30)/	/1-VR(10,60)/	/AR(10)/	/AR(30)/	Vs/Vp
Mean	33.66	44.99	16.84	29.53	4.25	4.73	0.39
z-Score	-20.12	-14.88	-5.27	-5.26			NA
t-Value					-0.95	-0.04	NA

5. Liquidity and Information Asymmetry

This paper follows the approach of Chordia, Roll and Subrahmanyam (2000) and Pukthuanthong-Le and Visaltanachoti (2009) to compute the effective spreads (EFFSPRD), absolute spreads (ASPRD), and relative spreads (RSPRD) as proxies for liquidity associated with informed trading. EFFSPRD is twice the absolute value of the difference between the trading price and the midpoint of the bid and ask quote occurring in the corresponding transaction. ASPRD is the absolute value of the difference between the bid and ask quotes occurring in the corresponding transaction. RSPRD is the absolute value of the difference between the bid and ask quote divided by the midpoint of the corresponding bid and ask quote. The first two measures are expressed in dollar value, and the last measure is expressed as a proportional value.

Table 3 shows the summary statistics for the spread measures. RSPRD has the lowest mean and standard deviation, while ASPRD has the highest mean and standard deviation. Moreover, since the effective spread is closer to actual transaction costs (Chordia et al., 2008), the ASPRD is likely to overestimate the transaction costs. This may be because the prices are likely to increase following a buy, and decrease following a sell (Madhavan, Richardson & Roomans, 1997). Large spreads lead to a low stock liquidity, since the associated high transaction costs prevent traders from participating in the market.

One approach of measuring liquidity without asymmetric information effects is based on the model of Glosten and Harris (1988). Glosten and Harris (1988) use a linear regression approach for estimating a model of the bid/ask spread. There are two components in this liquidity model: The adverse selection component (asymmetric information); and the transitory component, consisting of inventory costs, clearing fees, and monopoly profits. Since we are only interested in decomposing information asymmetry from liquidity, however, we simplify the model into a linear regression model that contains only two variables: Liquidity; and information asymmetry. The three spread measures computed above are used as proxies for the liquidity. The adjPIN is used as a measure of pure information asymmetry, as in the model of Duarte and Young (2009). In our simplified liquidity model, we use the residual value, or an error series { e }, to capture all of the other effects on liquidity not driven by information asymmetry.

Table 3: Summary Statistics of the Spread Measures

Table 3 shows the mean, standard deviation, 5th percentile, 25th percentile , median, 75th percentile, and 95th percentile of the cross-sectional series of spread measures for each firm, based on 1,044 firms during the period of January 3, 2007 to March 28, 2007. Intraday data is obtained from the CRSP and Reuters databases. RSPRD is the average relative bid-ask spread. ASPRD is the average absolute bid-ask spread. EFFSPRD is the average effective bid-ask spread. EFFSPRD and ASPRD are measured in dollar values. RSPRD is expressed as a percentage.

Variable	Mean	STD (%)	5th Pctl	25th Pctl	Median	75th Pctl	95th Pctl
EFFSPRD*100	2.60	4.84	1.05	1.39	1.85	2.67	5.53
ASPRD*100	3.57	6.83	1.32	1.92	2.58	3.78	7.83
RSPRD*100	0.12	0.13	0.03	0.05	0.08	0.13	0.32

Another measure for liquidity unrelated to information asymmetry is the probability of a symmetric order-flow shock, which can be observed directly from Duarte and Young's (2009) model. They decompose the probability of informed trading (PIN) into two separate components: Information asymmetry; and illiquidity. They then adopt adjusted PIN (adjPIN) as a proxy for the information asymmetry and the probability of a symmetric order-flow shock (PSOS) as a measure of illiquidity that is not related to information asymmetry, in their extended PIN model.

Duarte and Young's (2009) extended model is based on Easley, Kiefer, O'Hara and Paperman's (1996) PIN model. This model includes both informed trading on private information, and noise trading with exogenous reasons. An uninformed liquidity provider is also included in the model. Uninformed liquidity providers need compensation for the likelihood of trading with informed traders through the bid-ask spread (Duarte & Young, 2009). The Easley, Kiefer, O'Hara and Paperman (1996) PIN model is:

$$PIN = \frac{a \times u}{a \times u + \varepsilon_s + \varepsilon_b} \tag{7}$$

where variable a is the probability that a private information event will arrive on each given day, d is the probability that private information arrives and informed investors gain positive signals of such information, while the rate of trading for an informed trader is defined as u, and ε_b and ε_s are the noise traders' buy orders and sell orders, respectively.

The intuition is that, on each day, nature determines whether or not private information exists. On days with positive private information, both informed traders and noise traders participate in the market as buyers. The total buy order flow is, therefore, ε_b + u. On the other hand, only noise traders will set sell orders with the arrival rate ε_s, when the signal of private information is positive. On a day with negative private information, both informed traders and noise traders participate in the market as sellers. The total sell order flow is ε_s + u. Only noise traders will trade as buyers at rate ε_b. On a day with no private information, only noise traders participate in the market in both buy and sell positions, with rates ε_b and ε_s, respectively (Duarte & Young, 2009). See Figure 1 for information on the original PIN structure.

The maximum likelihood of calculating the parameters of PIN is:

$$L(\theta|B,S) = (1-a)e^{-\varepsilon_b}\frac{{\varepsilon_b}^B}{B!}e^{-\varepsilon_s}\frac{{\varepsilon_s}^S}{S!} + ade^{-(u+\varepsilon_b)}\frac{(u+\varepsilon_b)^B}{B!}e^{-\varepsilon_s}\frac{{\varepsilon_s}^S}{S!} + a(1-d)e^{-\varepsilon_b}\frac{{\varepsilon_b}^B}{B!}e^{-(u+\varepsilon_s)}\frac{(u+\varepsilon_s)^S}{S!} \tag{8}$$

Note that B and S are the number of buys and sells on a certain day, and $\theta = (a, u, \varepsilon_b, \varepsilon_s, d)$ is the parameter vector.

The model has two problems, however. First, the model does not capture the positive correlation between buy orders and sell orders. Second, the model does not show a large variance of buy and sell orders, compared with the actual variance of market trading activities (Duarte & Young, 2009).

This paper, therefore, uses Duarte and Young's (2009) extended PIN model, instead of the original PIN model because: The extended PIN model allows for simultaneous symmetric shocks of the buy and sell order flow; and, the extended PIN model allows for large buy and sell volatility. These two assumptions are also consistent with the empirical findings shown in Table 1. Therefore, this model is more appropriate for the analysis of real trading activities (Duarte & Young, 2009). Adding the effect of symmetric order-flow shocks to the model is important because disagreement between trades on public information may cause higher rates on both buy and sell trades (Duarte & Young, 2009). Different interpretations of public information among traders (Harris & Raviv, 1993; Kandel & Pearson, 1995), or diverse information signals observed by investors (He & Wang, 1995) are triggers of symmetric order flows. On the other hand, traders may purely participate in the market on particular days in order to reduce trading costs (Admati & Pfleiderer, 1988).

In Duarte and Young's (2009) model, the additional variable θ stands for the probability of symmetric order-flow shocks when no private information occurs, while θ' stands for the probability of symmetric order-flow shocks on days with private information. Δb and Δs stand for the additional arrival rate of buys and sells, respectively, when symmetric order-flow shocks occur (see Figure 2 for information on the extended PIN structure).

Figure 1: Tree Information for Easley, Kiefer, O'Hara and Paperman's (1996) original PIN Model

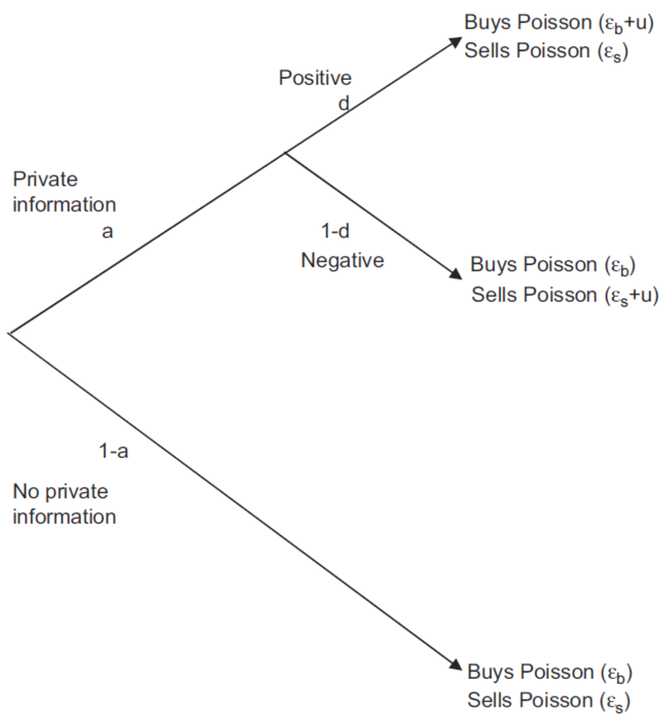
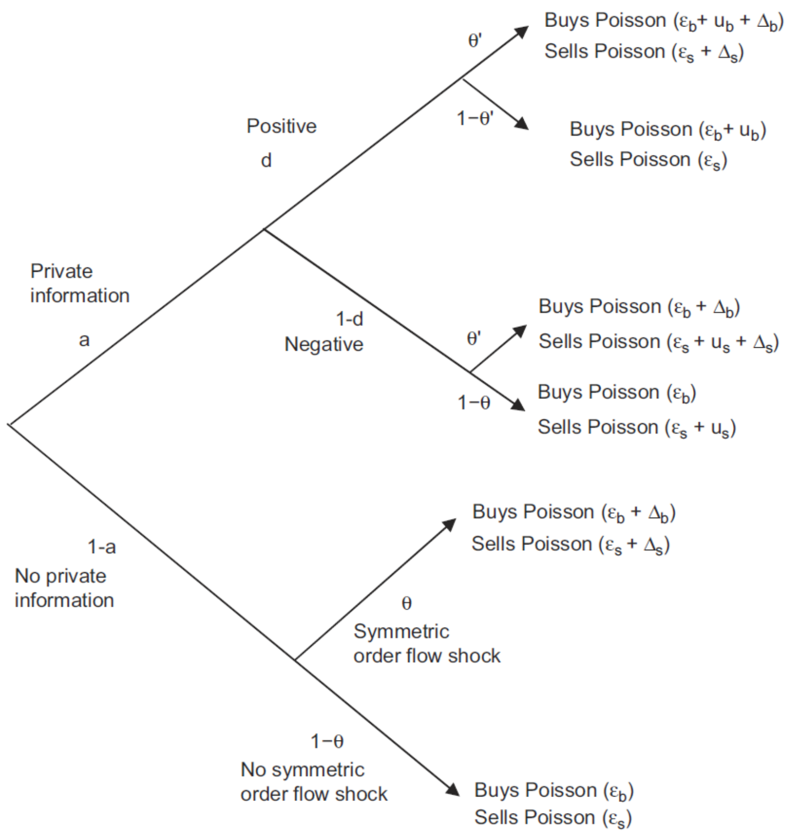


Figure 2: Tree Information for Duarte and Young's (2009) Extended PIN Model



The equations for adjPIN and PSOS are:

$$adjPIN = \frac{a \times (d \times u_d + (1-d) \times u_s)}{a \times (d \times u_d + (1-d) \times u_s) + (\Delta_b + \Delta_s) \times (a \times \theta' + (1-\mathrm{a}) \times \theta') + \varepsilon_s + \varepsilon_b} \tag{9}$$

$$PSOS = \frac{(\Delta_b + \Delta_s) \times (a \times \theta' + (1-\mathrm{a}) \times \theta)}{a \times (d \times u_d + (1-d) \times u_s) + (\Delta_b + \Delta_s) \times (a \times \theta' + (1-\mathrm{a}) \times \theta') + \varepsilon_s + \varepsilon_b}$$

(10)

The estimated parameters of adjPIN and PSOS are obtained from Duarte and Young's (2009) maximum likelihood function:

$$
\begin{aligned}
L(\theta|B,S) = {} & (1-a)(1-\theta)e^{-\varepsilon_b}\frac{{\varepsilon_b}^B}{B!}e^{-\varepsilon_s}\frac{{\varepsilon_s}^S}{S!} \\
& + (1-a)\theta e^{-(\varepsilon_b+\Delta_b)}\frac{(\varepsilon_b+\Delta_b)^B}{B!}e^{-(\varepsilon_s+\Delta_s)}\frac{(\varepsilon_s+\Delta_s)^S}{S!} \\
& + a(1-\theta')(1-d)e^{-\varepsilon_b}\frac{{\varepsilon_b}^B}{B!}e^{-(u_s+\varepsilon_s)}\frac{(u_s+\varepsilon_s)^S}{S!} \\
& + a\theta'(1-d)e^{-(\varepsilon_b+\Delta_b)}\frac{(\varepsilon_b+\Delta_b)^B}{B!}e^{-(u_s+\varepsilon_s+\Delta_s)}\frac{(u_s+\varepsilon_s+\Delta_s)^S}{S!} \\
& + a(1-\theta')de^{-(u_b+\varepsilon_b)}\frac{(u_b+\varepsilon_b)^B}{B!}e^{-\varepsilon_s}\frac{{\varepsilon_s}^S}{S!} \\
& + a\theta de^{-(u_b+\varepsilon_b+\Delta_b)}\frac{(u_b+\varepsilon_b+\Delta_b)^B}{B!}e^{-(\varepsilon_s+\Delta_s)}\frac{(\varepsilon_s+\Delta_s)^S}{S!}
\end{aligned}
$$

(11)

Note that B and S are the number of buys and sells on a particular day, while $\theta = (a, u_b, u_s, \varepsilon_b, \varepsilon_s, d, \theta, \theta', \Delta_b, \Delta_s)$ is the parameter vector.

Table 4 shows that both adjPIN and PSOS are significant in all percentiles. Also, we find that PSOS is higher than adjPIN in all percentiles, with the difference between PSOS and adjPIN becoming gradually larger, from 0.04 in the 5th percentile, to 0.13 in the 95th percentile. This result indicates that pure liquidity is more closely related to trading on any given day compared with asymmetry information, which is consistent with Duarte and Young's (2009) findings. Moreover, Table 4 also shows that the arrival rates of buy orders are higher than the rates of sell orders in all percentiles, for both informed traders and noise traders, which is consistent with the findings shown in Table 1.

Note that the model allows symmetric order-flow shocks on all days, but sets the restriction that the probability of symmetric order-flows is the same on days with and without private information ($\theta = \theta'$) (see Duarte and Young, 2009).

Table 5 shows the cross-sectional multivariate correlations between the spread measures adjPIN and PSOS. The spread measures are significantly consistent with each other. The adjPIN is positively related to the spread measures, with p-values all less than 0.0001. This is because asymmetric information is a component of liquidity (Campbell et al., 1997; Glosten & Harris, 1988). An increase in the amount of asymmetric information causes a decrease in liquidity. PSOS is also found to be positively related to the spread measures. This evidence supports Duarte and Young's (2009) argument that PSOS is a proxy for illiquidity unrelated to the information asymmetric effect. Moreover, there is a significant positive relation between adjPIN and PSOS, for which the coefficient is 0.353 (p-value<0.0001). Overall, the results indicate that an increase in private information causes an increase in adverse selection costs and, hence, a decrease in liquidity trading. This is due to the fact that liquidity traders need compensation to bear the risk of trading with informed traders.

Given that asymmetric information is a component of spread, the next step is to decompose this effect from spread and obtain the pure liquidity. As discussed earlier in this section, this paper runs the cross-sectional regression between the spread measures and adjPIN for each company, and then, use residual value or errors {e} to capture all other effects on liquidity that are not driven by information asymmetry.

Table 6 shows the cross-sectional regression between liquidity measures and adjPIN. There is strong evidence that adjPIN has a significant explanatory power on liquidity. The lowest standard error and highest R squares appear in the regression between the relative bid-ask spread and adjPIN, at 0.11 and 0.26, respectively. The highest standard error is found when using the absolute bid-ask spread as a proxy for illiquidity. The adjPIN has a similar explanatory power on the absolute bid-ask spread and effective bid-ask spread, of 2%.

In short, the evidence shows that there is a significantly positive relation among private information, the spreads, and pure liquidity unrelated to asymmetric information. An increase in the amount of private information leads to wider spreads (and higher PSOS) and, hence, a decrease in the stock's liquidity. The PSOS, like the residual error, captures the effects on liquidity without private information. Those effects may include inventory costs, order-processing costs/clearing fees, monopoly profits (Campbell et al., 1997; Glosten & Harris, 1988), and all other effects other than asymmetric information.

Table 4: Statistical Summary of the Estimation Results Based on Duarte and Young's (2009) Extended Model

Table 4 shows the cross-sectional estimated parameters based on Duarte and Young's (2009) extended model. The computation is based on 1,044 firms listed in the NYSE between January 3, 2007 and March 28, 2007. Intraday data is obtained from the CRSP and Reuters databases. The buys and sells initiated are identified using the approach developed by Lee and Ready (1991).

Variable a is the probability that a private information events will arrive on each given day; d is the probability that private information arrives and an informed investor gets a positive signal of such information; the rate of trading for informed traders to buy and sell are u_b and u_s, respectively; and ε_b and ε_s are noise traders' buy and sell orders, respectively. Variable θ stands for the probability of symmetric order-flow shocks when no private information occurs; θ' stands for the probability of symmetric order-flow shocks on the day with private information; and Δb andΔs stands for the additional arrival rate of buys and sells, respectively, when symmetric order-flow shocks occur. adjPIN is the estimated probability of informed trading in the extended model. PSOS is the estimated probability of symmetric order-flow shocks. The t-statistics for adjPIN and PSOS are in parentheses.

Note that a restriction of θ' = θ has been set according to Duarte and Young's (2009) preferred extended model.

Variable	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile
a	0.29	0.42	0.48	0.54	0.64
d	0.04	0.31	0.51	0.71	1.00
u_b	44.00	174.52	333.06	546.60	1267.41
u_s	34.26	146.93	272.94	490.85	1803.04
e_b	49.30	255.57	527.93	998.58	2488.71
e_s	37.47	223.33	430.95	887.90	2374.76
Δb	58.54	233.14	413.44	786.91	1755.30
Δs	59.59	224.20	408.93	716.85	1704.66
θ	0.14	0.24	0.31	0.39	0.52
adjPIN	0.06	0.08	0.10	0.12	0.18
t-Value	(4.39)	(6.20)	(7.22)	(8.21)	(10.11)
PSOS	0.10	0.15	0.18	0.23	0.31
t-Value	(3.32)	(4.94)	(6.20)	(7.70)	(10.22)

Table 5: Summary Statistics on the Multivariate Correlations Between the Spread Measures, adjPIN, and PSOS.

Table 5 shows the cross-sectional multivariate correlations between the spread measures, adjPIN, and PSOS. The computation is based on 1,044 firms listed in the NYSE between January 3, 2007 and March 28, 2007. Intraday data is obtained from the CRSP and Reuters databases. RSPRD is the average relative bid-ask spread. ASPRD is the average absolute bid-ask spread. EFFSPRD is the average effective bid-ask spread. The adjPIN is the estimated probability of informed trading in the extended model. PSOS is the estimated probability of symmetric order-flow shocks. adjPIN and PSOS are computed based on Duarte and Young's (2009) Extended PIN model. The p-values are in parentheses.

	EFFSPRD	**ASPRD**	**RSPRD**	**adjPIN**	**PSOS**
EFFSPRD	1.0000	0.997 (<.0001)	0.263 (<.0001)	0.154 (<.0001)	0.134 (<.0001)
ASPRD		1.0000	0.249 (<.0001)	0.150 (<.0001)	0.131 (<.0001)
RSPRD			1.0000	0.513 (<.0001)	0.318 (<.0001)
adjPIN				1.0000	0.353 (<.0001)
PSOS					1.0000

Table 6: Summary Statistics on the Cross-sectional Regression Between Spread Measures and adjPIN

This table presents the cross-sectional regression between the liquidity measures and adjPIN, using 1,044 NYSE-listed firms from January 3, 2007 to March 28, 2007. Intraday data is obtained from the CRSP and Reuters databases.

Table 6 shows the cross-sectional regression between the liquidity measures and the asymmetric information effects. RSPRD is the average relative bid-ask spread. ASPRD is the average absolute bid-ask spread. EFFSPRD is the average effective bid-ask spread. The dependent variable is adjPIN. The adjPIN is the estimated probability of informed trading in the extended model. adjPIN is computed based on Duarte and Young's (2009) Extended PIN model. The t-statistics are in parentheses.

Correction for heteroskedasticity has been considered in order to obtain consistent standard errors and covariance of the residues among the cross-sectional regressions in this sample. The t-statistics are in parentheses.

	Independent Variables								
	EFFSPRD			**ASPRD**			**RSPRD**		
Dependent Variables	Coef.	R^2	Std. Error	Coef.	R^2	Std. Error	Coef.	R^2	Std. Error
Intercept	0.77			1.06			-0.04		
	(2.10)			(2.09)			(-1.88)		
adjPIN	0.17	0.02	4.78	0.24	0.02	6.76	0.02	0.26	0.11
	(3.71)			(3.65)			(6.42)		

6. Cross-sectional Regression Between Pricing Efficiency and Liquidity

The previous section discussed the positive relation between the information asymmetry effect and liquidity. This section turns to examining the main hypothesis of this research: There is a significant and positive relation between pricing efficiency and liquidity, even after controlling for informed trading.

Table 7 shows the cross-sectional regression between the various measures of pricing efficiency and liquidity with informed trading risks. Correction of heteroskedasticity has been considered in order to obtain consistent standard errors and covariance of the residues among the cross-sectional regressions in this sample. Not surprisingly, a positive relation is found between price informational efficiency and the liquidity measures. EFFSPRD and ASPRD have significant explanatory power on pricing efficiency. RSPRD is found to be significantly related to pricing efficiency measured by the normalised pricing error and the thirty-minute midpoint return autocorrelation. Moreover, the impact of liquidity on pricing inefficiency is found to be more significant over a shorter time interval. In the regression between the variance ratio and EFFSPRD, the coefficients of the explanatory variable EFFSPRD decrease from 0.32 and 0.35 in the five-minute interval, to 0.24 and 0.29 in the ten-minute interval, respectively. In the regression between autocorrelation and EFFSPRD, the coefficients of the explanatory variable EFFSPRD decrease from 0.07 in the ten-minute interval, to 0.04 in the thirty-minute interval. The results from the other illiquidity measures are similar to the EFFSPRD measures, with the exception of a negative, but insignificant impact of RSPRD on /1-VR(10,60)/.

Overall, there is strong evidence that liquidity associated with informed trading effects has a significantly positive impact on pricing efficiency. Increases in spread (illiquidity) cause stock prices to move further away from a random walk. A stock with low liquidity is priced less efficiently than one with high liquidity. Such findings are inconsistent with the empirical results of Kyle (1984, 1985) and Subrahmanyam (1991), that increasing liquidity leaves the market efficiency unchanged, or even reduces pricing efficiency. This may be because the positive effects brought by increasing liquidity are more than offset by the adverse selection costs resulting from the entry of more informed traders into the market.

Table 8 shows the cross-sectional regression between pricing efficiency and liquidity after controlling for the asymmetric information effects. Correction for heteroskedasticity has also been considered here. The results indicate that pure liquidity unrelated to informed trading also has significant explanatory power on pricing efficiency. There are significantly positive impacts of e_{ASPRD} and $e_{EFFSPRD}$ on the degree of efficient pricing. The impact of pure liquidity unrelated to information asymmetry on pricing inefficiency is found to be more significant over a shorter time interval. An anomaly is found when using e_{RSPRD} as an explanatory variable for pricing efficiency. The series of residual e_{RSPRD} have a negative relation with all of the variance ratio measures. The t-test shows, however, that such a negative relation is not significant. When using PSOS as a proxy for pure liquidity, evidence that pure liquidity has a significant positive impact on pricing efficiency is also found.

Overall, an increase in pure liquidity trading can also improve the degree of pricing efficiency, as well as informed trading. This may be due to institutional trading activities. Boehmer, Kelley and Hall (2009) find that institutional holding and trading activities enhance price efficiency, even when they trade without superior information, if they are able to differentiate between permanent change (due to the revelation of new information) in stock prices and temporary deviations (pricing inefficiency due to the trades for exogenous reasons) from a stock's fundamental value. What is more, order processing itself can contribute to the degree of pricing efficiency. Learning from order flows in intraday trading, sophisticated investors can detect order imbalance, and participate in the market as counter parties to remove the serial dependence (Chordia et al., 2005).

Table 7: Summary Statistics on the Cross-sectional Regression Between Pricing Efficiency and Liquidity Combined with Informed Trading

This table presents the cross-sectional regression between the pricing efficiency measures and the liquidity measures, using 1,044 NYSE listed firms from January 3,2007 to March 28, 2007. Intraday data is obtained from the CRSP and Reuters databases.

Table 7 shows the cross-sectional regression between pricing efficiency and liquidity with asymmetric information effects. RSPRD is the average relative bid-ask spread. ASPRD is the average absolute bid-ask spread. EFFSPRD is the average effective bid-ask spread. The dependent variables are inverse indicators of pricing efficiency; Vs/Vp, /1-(VR5,30)/, /1-(VR5,60)/, /1-(VR10,30)/, /1-(VR10,60)/, /AR10/, and /AR30/.

Correction for heteroskedasticity has been considered, in order to have consistent standard errors and covariance of the residues across the cross-sectional regressions in this sample. The t-statistics are in parentheses.

Explanatory Variables	**Dependent Variables**						
	/1-(VR5,30)/	**/1-(VR5,60)/**	**/1-(VR10,30)/**	**/1-(VR10,60)/**	**/AR10/**	**/AR30/**	**Vs/Vp**
Intercept	32.82	44.09	16.22	28.78	4.06	4.63	0.35
	(42.64)	(58.26)	(43.35)	(65.76)	(34.77)	(37.69)	(16.79)
EFFSPRD	0.32	0.35	0.24	0.29	0.07	0.04	0.01
	(2.33)	(2.63)	(3.97)	(4.59)	(4.14)	(2.35)	(1.58)
Intercept	32.86	44.13	16.23	28.80	4.06	4.64	0.36
	(43.15)	(58.91)	(43.73)	(66.27)	(34.88)	(38.03)	(17.39)
ASPRD	0.22	0.24	0.17	0.20	0.05	0.03	0.01
	(2.38)	(2.69)	(4.13)	(4.80)	(4.51)	(2.41)	(1.66)
Intercept	33.12	44.69	16.78	29.60	4.21	4.34	0.06
	(33.35)	(45.97)	(35.20)	(52.69)	(27.31)	(23.87)	(0.48)
RSPRD	4.64	2.64	0.47	-0.59	0.42	3.33	2.82
	(0.69)	(0.40)	(0.15)	(-0.16)	(0.39)	(2.33)	(2.50)

Table 8: Summary Statistics on the Cross-sectional Regression Between Pricing Efficiency and Pure Liquidity

This table presents the cross-sectional regression between the pricing efficiency measures and the pure liquidity measures, using 1,044 NYSE listed firms from January 3, 2007 to March 28, 2007. Intraday data is obtained from the CRSP and Reuters databases.

Table 8 shows the cross-sectional regression between pricing efficiency and liquidity without asymmetric information effects. e_{RSPRD}, e_{ASPRD}, and $e_{EFFSPRD}$ are the residue errors computed from the cross-sectional regression between the spreads and adjPIN. PSOS is an alternative proxy for pure liquidity unrelated to information asymmetric effects, based on Duarte and Young (2009). The dependent variables are inverse indicators of pricing efficiency; Vs/Vp, /1-(VR5,30)/, /1-(VR5,60)/, /1-(VR10,30)/, /1-(VR10,60)/, /AR10/, and /AR30/.

Correction for heteroskedasticity has been considered in order to have consistent standard errors and covariance of the residues across the cross-sectional regressions in this sample. The t-statistics are in parentheses.

Explanatory Variables	Dependent Variables						
	/1-(VR5,30)/	**/1-(VR5,60)/**	**/1-(VR10,30)/**	**/1-(VR10,60)/**	**/AR10/**	**/AR30/**	**Vs/Vp**
Intercept	33.66	44.99	16.84	29.53	4.25	4.73	0.39
	(47.95)	(64.83)	(46.86)	(69.84)	(38.29)	(39.99)	(16.26)
E_EFFSPRD	0.27	0.31	0.23	0.29	0.07	0.03	0.01
	(2.41)	(2.77)	(4.05)	(4.62)	(4.34)	(2.25)	(1.50)
Intercept	33.66	44.99	16.84	29.53	4.25	4.73	0.39
	(47.95)	(64.83)	(46.86)	(69.84)	(38.29)	(39.99)	(16.26)
E_ASPRD	0.19	0.22	0.17	0.21	0.05	0.02	0.00
	(2.48)	(2.84)	(4.22)	(4.83)	(4.72)	(2.27)	(1.53)
Intercept	33.66	44.99	16.84	29.53	4.25	4.73	0.39
	(47.88)	(64.71)	(46.64)	(69.49)	(38.10)	(40.07)	(17.60)
E_RSPRD	-3.32	-4.56	-0.94	-1.60	0.44	2.54	2.72
	(-0.45)	(-0.64)	(-0.28)	(-0.41)	(0.38)	(1.73)	(2.05)
Intercept	28.86	41.04	15.42	28.66	4.75	4.10	0.26
	(14.37)	(20.69)	(15.62)	(23.89)	(15.08)	(11.80)	(3.94)
PSOS	0.25	0.21	0.07	0.05	-0.03	0.03	0.01
	(2.45)	(2.03)	(1.39)	(0.71)	(-1.71)	(1.83)	(1.67)

7. Limitations

There are several limitations to this paper. First, the quoted bid-ask spreads may be noisy measures of liquidity. This is because a large number of trades occur outside the spread, and there are also a small number of trades occurring within the spread (Brennan & Subrahmanyama, 1996; Lee, 1993). Second, computation precision depends on accuracy in indentifying buyer- and seller- initiated trades. Ellis, Michaely and O'Hara (2000) find that only 81.05% of the trade sign can be accurately classified using the approach developed by Lee and Ready (1991). The bias in the accuracy of identifying trades directly contributes to potential misleads in our analysis. Third, it is impossible to capture the pricing efficiency using any single measure. According to Boehmer et al. (2009), returns autocorrelation and variance ratios may not be accurately represent price informational efficiency. These measures do not distinguish between permanent and temporary price changes. For instance, if informed traders split their orders over time, and prices gradually approach their fundamentals, this will cause positive autocorrelation and large divergence from 1 in the variance ratios, although all of the publicly available information is well processed. Also, autocorrelations and variance ratios disregard variation in transaction prices, because they are calculated in clock time. Furthermore, it is necessary to compute the unobserved random walk component in order to calculate pricing errors. This imposes a problem when infrequently traded stocks do not follow a random walk (see Boehmer et al., 2009; Fama & French, 1988; Hasbrouck, 1993; Lo & MacKinlay, 1988; Poterba, Lawrence & Summers, 1988). Also, if it takes too long to converge the temporary deviation to an efficient price, "…the variance decomposition will erroneously attribute deviations to changes in the efficient price" (Boehmer et al., 2009, p. 9).

Taking these issues into careful consideration, this paper does not focus on testing the pricing efficiency, or the liquidity, in an absolute sense. Rather, these techniques are used together as comprehensive instruments, or indices, to explore the relation between relative pricing efficiency and the explanatory variables.

8. Conclusion

This paper tests the relation between pricing efficiency and liquidity with, and without, the effect of private information. First, we show that informed trading is negatively related to the level of liquidity. This result is consistent with previous researches that find that informed trading reduces liquidity. Second, this report provides a way to explore the direct relation between price efficiency and liquidity by applying cross-sectional regressions. Empirical evidences support the hypothesis that liquidity associated with asymmetric information effects enhances pricing efficiency. The cross-sectional analysis between relative informational efficiency and liquidity, combined with informed trading, indicates that liquidity does have explanatory power on pricing efficiency. Third, since the motivation of implementing insiders' trading rules, and reducing the minimum tick size, is to improve liquidity trading, we test whether pure liquidity trading has an impact on pricing efficiency. We find that pure liquidity trading also contributes to price informativeness. This may be because liquidity trading includes institutional trading activities. Institutional holding and trading activities can facilitate informational efficiency, even without access to private information, when it is possible for traders to differentiate between permanent change and temporary change in stock prices (Boehmer et al., 2009). Moreover, order processing itself can contribute to the degree of pricing efficiency. Learning from order flows within the trading day, sophisticated investors can detect order imbalance, and participate in the market as counter parties to remove the serial dependence (Chordia et al., 2005). Indeed, the reasons why pure liquidity trading has a positive impact on the degree of pricing efficiency requires further study.

Appendices

SAS Codes

```
OPTIONS compress=yes ;
libname a 'H:\nyseJanMar2007 working file\data_pt1 analysis';
proc sort data=a.data_pt1;
by ric date time type;
run;

*******************************************************************

Combine all trades at the same time and price into one;

*******************************************************************;

**create data set trades;

data a.trades;
set a.data_pt1 (where=(type = 'T'));

if price<=0 then delete;
if price= . then delete;

if volume=0 then delete;
if volume= . then delete;

if time<'9:30:00't then delete;
if time>'16:00:00't then delete;

volume=volume/1000;*\Note: inorder to avoid scaling problem, I divide volume by
1000\;
drop BidPrice BidSize AskPrice AskSize Qualifiers;
run;

proc sort data=a.trades out=trades;
by ric date time price;
run;

proc means data=trades noprint;
by ric date time price;
output out=adjtrades (rename=(_freq_=numtrades) drop =_type_)
sum (volume) = volume;
run;
*********************************************************************

1.0 Compute tick test based on trade prices and various spread measures;
```

```
*********************************************************************

1.1 tick test;

*********************************************************************;

data ntrades;
set adjtrades;
tid=_n_; *\create unique trade identifier\;
lagprice=lag(price);
lag2price=lag2(price);

if price > lagprice then tick = 1;
if price < lagprice then tick = -1;

if price = lagprice then do;
  if lagprice>lag2price then tick =1;
  if lagprice<lag2price then tick =-1;
  end;
  if _n_<3 then tick=0;
  if tick= . then tick=0;
  drop lagprice lag2price;
  label tick = 'trade indicator based on tick test';
  label tid = 'trade identifier';
  label numtrades = 'number of aggregated trades';
  run;
***Frequency analysis for tick test;

*proc freq data=ntrades;
  *by ric;
  *table tick;
  *run;

*****************************************************************

1.2 Compute quotes change and combine them with trade records;

*****************************************************************;

**create date set quotes;

data a.quotes;
set a.data_pt1 (where=(type = 'Q'));
if bidprice<=0 then delete;
if bidprice= . then delete;
if bidsize<=0 then delete;
if bidsize= . then delete;

if askprice<=0 then delete;
if askprice= . then delete;
```

```
if asksize<=0 then delete;
if asksize=. then delete;

if time<'9:30:00't then delete;
if time>'16:00:00't then delete;

drop Price Volume VWAP Qualifiers;
run;

**compute quote changes;

proc sort data=a.quotes out=quotes;
by ric date time;
run;

data allqchange;
set quotes;
by ric;
midpoint=(BidPrice+AskPrice)/2;
oldmp=lag(midpoint);
if first.ric then oldmp= .;
qid=_n_;*\create unique quote identifier\;
drop oldmp;
label qid ='quote identifier';
label midpoint ='quote midpoint';
if midpoint ne oldmp then output;
run;
**combine trades and quotes;

data qandt;
set allqchange(in=a) ntrades(in=b);
if a then trade=0;
if b then trade=1;
run;
************************************************************************
**********************************

1.3 Compute various spread measures, tradesign, signed_trading_volume,and
sqrt_signed_trading_volume;

************************************************************************
*********************************;

title1 'Spread estimation and trade direction';

proc sort data = qandt;
by ric date time;
run;

data a.spread;
```

```
set qandt;
by ric date;

*reset retained variable if a new ticker or new day starts;
if first.ric or first.date then do;
   nbid= .; nofr= .; currentmidpoint= .;
   end;

*assign bid and ask to new variables for retaining;
if BidPrice ne . then nbid=BidPrice;
if AskPrice ne . then nofr=AskPrice;
if midpoint ne . then currentmidpoint = midpoint;

*compute spread measures;
effsprd = abs(Price-(nbid+nofr)/2)*2;
asprd = nofr - nbid;
rsprd = asprd/((nbid+nofr)/2);

*compute tradesign;
if currentmidpoint ne . then do;
* test based on quotes - compare current trade to quote: -1 is sell initiated, 1 is buy
initiated;

  if price > currentmidpoint then tradesign=1;
  if price < currentmidpoint then tradesign=-1;
  if price = currentmidpoint then do;
  *tick size test;
     if tick = 1 then tradesign =1;
           if tick = -1 then tradesign =-1;
           if tick = 0 then tradesign = 0;
  end;

*compute signed trading volume;
signed_trading_volume= tradesign * volume; *\ in shares\;

*compute signed square root of trading volume;
if signed_trading_volume >0 then sqrt_signed_trading_volume =
Sqrt(signed_trading_volume);
if signed_trading_volume <0 then sqrt_signed_trading_volume =
(Sqrt(abs(signed_trading_volume ) ))*(-1);
end;

label nbid = 'last outstanding bid';
label nofr = 'last outstand ofr';
label effsprd = 'effective spread';
label asprd = 'absolute spread';
label rsprd = 'relative spread';
label tradesign ='indicator for trade direction';
label signed_trading_volume ='signed trading volume';
label sqrt_signed_trading_volume ='signed square root of trading volume';
```

```
if trade = 1 then output a.spread;
retain nbid nofr currentmidpoint;
drop BidPrice AskPrice midpoint qid trade;
run;
***Compute descriptive statistics for tradesign, signed trading volume, signed square
root of trading volume and spread measures;

*proc means data = a.spread n mean median min max ;
*by ric;
*var price volume effsprd asprd rsprd tradesign signed_trading_volume
sqrt_signed_trading_volume;
*run;
**********************************************************************
********

2.0 Pricing Error Estimation

**********************************************************************
********;

**********************************************************************
********

2.1 Preparation for VAR estimation: resort trades and quotes in reverse time;

**********************************************************************
********;

*sort backwards in time to match trades to subsequent quote changes;

title1 'VAR estimaion';

proc sort data = qandt;

by ric descending date descending time descending qid;
run;
***Find quote updates associated with prior trades;

*associate order flow with quote changes in trade time;

data tradematch;
set qandt;
by ric descending date;
lagtrade=lag(trade);
lagric=lag(ric);

*reset retained variables if a new ticker or new day strats;
if first.ric and first.date then do;
end;
```

```
*quote resords;
*assign bid and ask to variables for retaining, if quote data is nonmissing;
if trade=0 then do;
nbid=BidPrice; nofr=AskPrice; nqid=qid; qtime=time; nmidpoint= midpoint;
end;

*trade records;
if trade=1 and lagtrade=0 and ric=lagric then do;
 qage=qtime-time;
 if qage=<0 then do;*means no delay problem;
 BidPrice=nbid;
 AskPrice=nofr;
 qid=nqid;
 midpoint=nmidpoint;
 end;
format qtime time8.;
end;
label BidPrice = 'last outstanding bid';
label AskPrice = 'last outstanding ask';
label qtime = ' time of last quote';
label qage= 'delay between trade and quote';
retain nbid nofr qtime nmidpoint nqid;
drop nmidpoint nbid nofr nqid lagric lagtrade trade;
*output;
if trade=1 then output tradematch;
run;
***Find quote updates not associated with trades;

*identify all unique quotes that were matched to a trade;

proc sort data=tradematch (keep=qid qage where=(qage<=0))
out = qids nodupkey;
by qid;
run;
*merge with all original quotes and retain only those not already matched with trades;

data quotematch;
merge allqchange (in=a) qids (in=b drop=qage);
by qid;
if not b;
run;
***Add unmatched quote updates to the trade-and-quote data set;

*combine unmatched quote with those already matched to trades;
*this adds one record for each unmatched quote;

data qnspread;
set tradematch (drop=BidPrice AskPrice  numtrades volume qtime) quotematch
(in=new);
unmatchedquote=new;
```

```
run;

proc sort data=qnspread;
by ric date time qid;
run;
*compute midpoint changes and trading price changes;
*convert trading price to logprice;*logprice changes in percentage;

data allqncspread;
set qnspread;
by ric;

price=log(price);*change trading price to log price;
midpoint=log(midpoint);*change midpoint to log midpoint;

if not first.ric then do;
   if midpoint ne . then currentmidpoint = midpoint;
   mpdf=(currentmidpoint - lag(currentmidpoint))*100;*midpoint difference in
percentage;
   if price ne . then  currentprice=price;
   pricedf=(currentprice-lag(currentprice))*100;*logprice difference in percentage;
   end;
retain currentmidpoint currentprice;
drop currentmidpoint currentprice;
run;
***Add trade direction variables to the trade-and-quote data and compute descriptive
statistics;

proc sort data=allqncspread;
by tid;
run;
**Create dataset for VAR estimation;

data a.vardata;
merge allqncspread (keep=ric date time tid unmatchedquote mpdf pricedf)
      a.spread(keep=tid signed_trading_volume tradesign tick
sqrt_signed_trading_volume);
by tid;
if unmatchedquote then do;
   signed_trading_volume=0;
   sqrt_signed_trading_volume=0;
   tick=0;
   tradesign=0;
   end;
label mpdf = 'change in logged quote midpoint (%)';
label pricedf = 'change in logged trading price (%)';
label tradesign ='indicator for trade direction';
label signed_trading_volume ='signed trading volume (000s) ';
label sqrt_signed_trading_volume ='signed square root of trading volume (000s)';
label tick = 'tick test';
```

```
drop unmatchedquote;
run;

proc sort data=a.vardata;
by ric date time;
run;

*proc means data=a.vardata n mean median min max;
*var mpdf pricedf  tick signed_trading_volume tradesign  sqrt_signed_trading_volume;
*by ric;
*run;
*********************************************************************
************

2.2 VAR Estimation;

*********************************************************************
*************
*important note: here we use transaction price (pret) instead of quote-midpoint return
(mpret) to estimate VAR, regarding to Hasbrouck 1993;
***VAR regression model;
*estimation by accumulated impulse response function;
*Estimate VAR based on 'pricedf' and 'signed_trading_volume' 'tradesign'  and
'sqrt_signed_trading_volume';

proc printto print="G:\nyseJanMar2007 working file\data.out"; run;
%let nImpulse=15;

proc varmax data=a.vardata (where = (pricedf ne . and tradesign ne . and
signed_trading_volume ne . and sqrt_signed_trading_volume ne . ));
   by ric;
      model tradesign, signed_trading_volume, sqrt_signed_trading_volume, pricedf /
p=5 dif=(pricedf(1)) noint print=(impulse(&nImpulse)=(accum ortho));
      ods output covInnovation=covInnovation;
      ods output accumImpulse=accumImpulse;
      ods output orthoImpulse=orthoImpulse;
      run;

*proc varmax data=a.vardata (where = (pricedf ne . and tradesign ne . and
signed_trading_volume ne . and sqrt_signed_trading_volume ne . ));
   *by ric;
      *model tradesign, signed_trading_volume, sqrt_signed_trading_volume, pricedf
/ p=5 dif=(pricedf(1)) noint print=(impulse(&nImpulse)=(accum ortho));
      *ods output covInnovation=covInnovation;
      *ods output accumImpulse=accumImpulse;
      *ods output orthoImpulse=orthoImpulse;
      *run;

proc printto; run;
```

```
data lastAccumImpulse;

        set accumImpulse;
        where lag=&nImpulse and variable='pricedf';
        drop lag;

run;

*data lastAccumImpulse;

        *set accumImpulse;
        *where lag=&nImpulse and variable='pricedf';
        *drop lag;
*run;


*proc print data=lastAccumImpulse;
*run;
*proc print data=CovInnovation;
*run;


*********************************************************************
*************

2.3 Compute Pricing Errors according to Hasbrouck 1993 and Boehmer and Kelly 2009;

*********************************************************************
*************

*2.3.1 Compute Pricing Error Variance ;


title1'Pricing error Estimation';

data indiv1;
set lastAccumImpulse;
by ric;
keep ric;
run;

data indiv2;
set lastAccumImpulse;
by ric;
drop ric;
run;

data indiv3;
set CovInnovation;
by ric;
```

```
drop ric;
run;

proc iml;
use indiv2;
read all  into x ;

use indiv3;
read all  into y ;

use indiv1;
read all var _char_  into names;

n=nrow(x);
m=ncol(y);
prv=j( n,1,0);


do i = 1 to n;
iystart = 4*(i-1)+1;
iyend = 4*(i-1)+4;
prv[i,]=x[i,]*y[iystart : iyend,]*T(x[i,]);

*print prv;
end;

*print prv;

create  Pricing_Error_Variance from prv [colname='Pricing_Error_Var'];
append from prv;
quit;

*proc print data=Pricing_Error_Variance;*run;

data a.pricing_error_variance;
merge indiv1 Pricing_Error_Variance;
run;


*2.3.2 Compute Normalized pricing error STD;


data logprice;
set trades;
by ric;
logprice=log(price);
run;

proc means data=logprice n mean std var noprint;
by ric;
```

```
var logprice;
output out=STD_logprice std=Vp;
run;

data a.normalized_pricing_error_STD;
merge a.pricing_error_variance  STD_logprice;
by ric;

Vs=sqrt(Pricing_Error_Var);

NormalisedPricingError=Vs/Vp; /*express in % */

*log_pricing_error=log(Vs);

label NormalisedPricingError ='Normalised pricing error STD';
label Vs ='Pricing error STD';
label Vp ='STD of logged trading price';

title2 'Normalised pricing error';
drop _type_ _freq_  ;
run;
**********************************************************************
****

3.0 Variance Ratio Estimation

**********************************************************************
****

3.1 Variance ratio (5min,30min) and (5min,60min)

**********************************************************************
****

3.1.1 Variance Ratio (5min, 30min)

**********************************************************************
****;


title1 'Variance ratio analysis';

***Construct five minute time interval;

data data2;
	set a.spread;
	by ric;
	DT=DHMS(Date,0,0,Time);
	format Date date9. Time time11.2 DT datetime20.3;
	if volume = . then delete;
```

```
        run;

proc sort data=data2;
        by ric dt;
        run;

proc means data=data2 noprint;
        class ric date;
        output out=tmp n=cnt;
        run;
* select only ric-date;

data date;
        set tmp;

        if _type_ ne 3 then delete;
        keep ric date;
        run;
*construct time interval;

data interval;
        informat start time11.2 end time11.2;
        input start end;
        format start time11.2 end time11.2;
        datalines;
9:30:00      9:35:00
9:35:00      9:40:00
9:40:00      9:45:00
9:45:00      9:50:00
9:50:00      9:55:00
9:55:00      10:00:00
10:00:00     10:05:00
10:05:00     10:10:00
10:10:00     10:15:00
10:15:00     10:20:00
10:20:00     10:25:00
10:25:00     10:30:00
10:30:00     10:35:00
10:35:00     10:40:00
10:40:00     10:45:00
10:45:00     10:50:00
10:50:00     10:55:00
10:55:00     11:00:00
11:00:00     11:05:00
11:05:00     11:10:00
11:10:00     11:15:00
11:15:00     11:20:00
11:20:00     11:25:00
11:25:00     11:30:00
11:30:00  11:35:00
```

11:35:00	11:40:00
11:40:00	11:45:00
11:45:00	11:50:00
11:50:00	11:55:00
11:55:00	12:00:00
12:00:00	12:05:00
12:05:00	12:10:00
12:10:00	12:15:00
12:15:00	12:20:00
12:20:00	12:25:00
12:25:00	12:30:00
12:30:00	12:35:00
12:35:00	12:40:00
12:40:00	12:45:00
12:45:00	12:50:00
12:50:00	12:55:00
12:55:00	13:00:00
13:00:00	13:05:00
13:05:00	13:10:00
13:10:00	13:15:00
13:15:00	13:20:00
13:20:00	13:25:00
13:25:00	13:30:00
13:30:00	13:35:00
13:35:00	13:40:00
13:40:00	13:45:00
13:45:00	13:50:00
13:50:00	13:55:00
13:55:00	14:00:00
14:00:00	14:05:00
14:05:00	14:10:00
14:10:00	14:15:00
14:15:00	14:20:00
14:20:00	14:25:00
14:25:00	14:30:00
14:30:00	14:35:00
14:35:00	14:40:00
14:40:00	14:45:00
14:45:00	14:50:00
14:50:00	14:55:00
14:55:00	15:00:00
15:00:00	15:05:00
15:05:00	15:10:00
15:10:00	15:15:00
15:15:00	15:20:00
15:20:00	15:25:00
15:25:00	15:30:00
15:30:00	15:35:00
15:35:00	15:40:00
15:40:00	15:45:00

```
15:45:00        15:50:00
15:50:00        15:55:00
15:55:00        16:00:00
;

* combine interval into ric-date;

proc sql;
  create table ricdi as
  select *
  from date, interval ;

quit;

proc sort data=ricdi;
by ric date start;
run;
* combine date & time;

data ricdti;
        set ricdi;

        dtstart = dhms(date,0,0,start);
        dtend = dhms(date,0,0,end);
        format dtstart datetime20.2 dtend datetime20.2;
        dt = dtend;
        keep ric date dtstart dtend dt;
        run;

proc sort data=ricdti;
        by ric dtstart;
        run;
**construct five-minute time interval data;

data interval;
        set ricdti data2;
        run;

proc sort data=interval;
        by ric dt;
        run;

data a.interval5min (keep=ric date dtstart dtend endprice endmidpoint db ds nb ns vb
vs);
        set interval;
        by ric;
        retain dtstart dtend db ds nb ns vb vs endprice endmidpoint;
        if _N_=1 then do; db=0; ds=0; nb=0; ns=0; vb=0; vs=0; end;

        if dtstart = . then
```

```
            do;
            if tradesign = 1 then
                do;
                db = db + price*volume;
                nb = nb + 1;
                vb = vb + volume;
                end;
            if tradesign = -1 then
                do;
                ds = ds + price*volume;
                ns = ns + 1;
                vs = vs + volume;
                end;
                endprice=price;
                endmidpoint=currentmidpoint;
            end;
        else
            do; output; db=0; ds=0; nb=0; ns=0; vb=0; vs=0; end;
        run;
*proc print data=interval5min(obs=100);*run;


***Variance Ratio (5min, 30min);

title2'Variance Ratio for 30-minute Returns to 5-minute returns';

proc sort data=a.interval5min;
by ric dtstart dtend;
run;

data mpret5min;
set a.interval5min;
by ric date;
retain n;

  mpret=log(endmidpoint / lag(endmidpoint));*/logged five minutes return;
  mpret6=log(endmidpoint / lag6(endmidpoint));*/logged thirty minutes return;
  mpret12=log(endmidpoint / lag12(endmidpoint));*/logged sixty minutes return;

if first.ric or first.date then do;
  n=0;
  end;
n+1;

if n<=1 then  mpret= .;
if n<=6 then mpret6= .;
if n<=12 then mpret12= .;

run;
*proc print data=mpret5min (obs=20);*run;
```

```
proc means data=mpret5min n mean std var noprint;
by ric;
var mpret;
output out=muhat mean=muhat n=nq;
run;

data mpret30min;
if _n_=1 then set muhat;
set mpret5min;
by ric;
sigatop=((mpret-muhat)**2);*/(P(k)-P(k-1)-muhat))^2;*/squared deviations from mean
return;

*/sum((P(k)-P(k-1)-muhat)^2 * (P(k-j)-P(k-j-1)-muhat)^2);*/product of current and
lagged squared deviation;
deltop1=sigatop*lag(sigatop);
deltop2=sigatop*lag2(sigatop);
deltop3=sigatop*lag3(sigatop);
deltop4=sigatop*lag4(sigatop);
deltop5=sigatop*lag5(sigatop);



delbot=sigatop;*/(P(k)-P(k-1)-muhat))^2;*/denominator used in estimation delta;
sigctop=((mpret6 - 6*muhat)**2);*/squared deviation from twice the mean return (for
thirty-minute return);
run;

proc means data=mpret30min noprint;
by ric;
var sigatop sigctop deltop1 deltop2 deltop3 deltop4 deltop5  delbot;
output out=varrat5min_30min
sum=sigatop sigctop deltop1 deltop2 deltop3 deltop4 deltop5  delbot;
run;

data a.varrat5min_30min;
set varrat5min_30min;
by ric;

q = 6;  qm1 = q - 1;
        j = 1;
        theta=0;
        array deltaj(5) deltop1 -- deltop5;
        array delta(5);
        nq = _freq_;
        m = q*(nq-q+1)*(1-q/nq);
        siga = sigatop/(nq-1);
        sigc = sigctop/m;
```

```
        varrat5min_30min = sigc/siga;
        do until (j > qm1);
                delta(j) = nq*deltaj(j)/(delbot**2);
                theta = theta + ((2*(q-j)/q)**2)*delta(j);
                j+1;
        end;

   invVR5min_30min=abs(1-varrat5min_30min);
        z = sqrt(nq)*(varrat5min_30min-1)/sqrt(theta);



        keep ric nq varrat5min_30min invVR5min_30min z;
        label
        nq = "Number of Weekly Returns"
        varrat5min_30min = "VR(5,30)"
        z = "Heteroskedastic Robust Test Statistic"
   invVR5min_30min="/1-VR(5,30)/";
        title2'Variance Ratio for 30-minute Returns to 5-minute returns';
        run;

*proc print data=varrat5min_30min label noobs;
        *run;
****************************************************

3.1.2 Variance Ratio (5min, 60min);

*****************************************************;

title2'Variance Ratio for 60-minute Returns to 5-minute returns';

data mpret60min;
if _n_=1 then set muhat;
set mpret5min;
by ric;
sigatop=((mpret-muhat)**2);*/(P(k)-P(k-1)-muhat))^2;*/squared deviations from mean return;

*/sum((P(k)-P(k-1)-muhat)^2 * (P(k-j)-P(k-j-1)-muhat)^2);*/product of current and lagged squared deviation;
deltop1=sigatop*lag(sigatop);
deltop2=sigatop*lag2(sigatop);
deltop3=sigatop*lag3(sigatop);
deltop4=sigatop*lag4(sigatop);
deltop5=sigatop*lag5(sigatop);
deltop6=sigatop*lag6(sigatop);
deltop7=sigatop*lag7(sigatop);
deltop8=sigatop*lag8(sigatop);
deltop9=sigatop*lag9(sigatop);
deltop10=sigatop*lag10(sigatop);
```

```
deltop11=sigatop*lag11(sigatop);


delbot=sigatop;*/(P(k)-P(k-1)-muhat))^2;*/denominator used in estimation delta;
sigctop=((mpret12-12*muhat)**2);*/squared deviation from twice the mean return (for
ten-minute return);
run;

proc means data=mpret60min noprint;
by ric;
var sigatop sigctop deltop1 deltop2 deltop3 deltop4 deltop5 deltop6
deltop7 deltop8 deltop9 deltop10 deltop11  delbot;
output out=varrat5min_60min
sum=sigatop sigctop deltop1 deltop2 deltop3 deltop4 deltop5 deltop6
deltop7 deltop8 deltop9 deltop10 deltop11  delbot;
run;

data a.varrat5min_60min;
set varrat5min_60min;
by ric;
q = 12;  qm1 = q - 1;
        j = 1;
        theta=0;
        array deltaj(11) deltop1 -- deltop11;
        array delta(11);
        nq = _freq_;
        m = q*(nq-q+1)*(1-q/nq);
        siga = sigatop/(nq-1);
        sigc = sigctop/m;
        varrat5min_60min = sigc/siga;
        do until (j > qm1);
                delta(j) = nq*deltaj(j)/(delbot**2);
                theta = theta + ((2*(q-j)/q)**2)*delta(j);
                j+1;
        end;

   invVR5min_60min=abs(1-varrat5min_60min);
        z = sqrt(nq)*(varrat5min_60min-1)/sqrt(theta);



        keep ric nq varrat5min_60min invVR5min_60min z;
        label
        nq = "Number of Weekly Returns"
        varrat5min_60min = "VR(5,60)"
        z = "Heteroskedastic Robust Test Statistic"
   invVR5min_60min="/1-VR(5,60)/";
        title2'Variance Ratio for 60-minute Returns to 5-minute returns';
        run;
```

```
*proc print data=varrat5min_60min label noobs;
	*run;
****************************************************************

3.2 Variance ratio (10min,30min) and (10min,60min)

****************************************************************

3.2.1 Variance ratio (10min,30min);

*****************************************************************

***Construct ten minute interval;



data interval2;
	informat start time11.2 end time11.2;
	input start end;
	format start time11.2 end time11.2;
	datalines;
9:30:00		9:40:00
9:40:00		9:50:00
9:50:00		10:00:00
10:00:00 10:10:00
10:10:00 10:20:00
10:20:00 10:30:00
10:30:00 10:40:00
10:40:00 10:50:00
10:50:00 11:00:00
11:00:00 11:10:00
11:10:00 11:20:00
11:20:00 11:30:00
11:30:00 11:40:00
11:40:00 11:50:00
11:50:00 12:00:00
12:00:00 12:10:00
12:10:00 12:20:00
12:20:00 12:30:00
12:30:00 12:40:00
12:40:00 12:50:00
12:50:00 13:00:00
13:00:00 13:10:00
13:10:00 13:20:00
13:20:00 13:30:00
13:30:00 13:40:00
13:40:00 13:50:00
13:50:00 14:00:00
14:00:00 14:10:00
14:10:00 14:20:00
```

```
14:20:00 14:30:00
14:30:00 14:40:00
14:40:00 14:50:00
14:50:00 15:00:00
15:00:00 15:10:00
15:10:00 15:20:00
15:20:00 15:30:00
15:30:00 15:40:00
15:40:00 15:50:00
15:50:00 16:00:00
;
proc sql;
  create table ricdi2 as
  select *
  from date, interval2 ;
quit;

proc sort data=ricdi2;
by ric date start;
run;
* combine date & time;

data ricdti2;
        set ricdi2;
        dtstart = dhms(date,0,0,start);
        dtend = dhms(date,0,0,end);
        format dtstart datetime20.2 dtend datetime20.2;
        dt = dtend;
        keep ric date dtstart dtend dt;
        run;

proc sort data=ricdti2;
        by ric dtstart;
        run;
**construct ten-minute time interval data;

data interval2;
        set ricdti2 data2;
        run;

proc sort data=interval2;
        by ric dt;
        run;

data a.interval10min (keep=ric date dtstart dtend endprice endmidpoint db ds nb ns vb
vs);
        set interval2;
        by ric;
        retain dtstart dtend db ds nb ns vb vs endprice endmidpoint;
        if _N_=1 then do; db=0; ds=0; nb=0; ns=0; vb=0; vs=0; end;
```

```
        if dtstart = . then
                do;
                if tradesign = 1 then
                        do;
                        db = db + price*volume;
                        nb = nb + 1;
                        vb = vb + volume;
                        end;
                if tradesign = -1 then
                        do;
                        ds = ds + price*volume;
                        ns = ns + 1;
                        vs = vs + volume;
                        end;
                        endprice=price;
                        endmidpoint=currentmidpoint;
                end;
        else
                do; output; db=0; ds=0; nb=0; ns=0; vb=0; vs=0; end;
        run;
*proc print data=a.interval10min (obs=100);*run;

**Variance Ratio(10min,30min);

title2'Variance Ratio for 30-minute Returns to 10-minute returns';
proc sort data=a.interval10min;
by ric dtstart dtend;
run;

data mpret10min2;
set a.interval10min;
by ric date;
retain n;

   mpret=log(endmidpoint / lag(endmidpoint));*/logged ten minutes return;
   mpret3=log(endmidpoint / lag3(endmidpoint));*/logged thirty minutes return;
   mpret6=log(endmidpoint / lag6(endmidpoint));*/logged sixty  minutes return;

if  first.ric or  first.date then do;
    n=0;
    end;
n+1;

if n<=1 then  mpret= .;
if n<=3 then mpret3= .;
if n<=6 then mpret6= .;


run;
```

```
*proc print data=mpret10min2 (obs=20);*run;


proc means data=mpret10min2 n mean std var noprint;
by ric;
var mpret;
output out=muhat mean=muhat n=nq;
run;

data mpret30min2;
if _n_=1 then set muhat;
set mpret10min2;
by ric;
sigatop=((mpret-muhat)**2);*/(P(k)-P(k-1)-muhat))^2;*/squared deviations from mean
return;

*/sum((P(k)-P(k-1)-muhat)^2 * (P(k-j)-P(k-j-1)-muhat)^2);*/product of current and
lagged squared deviation;
deltop1=sigatop*lag(sigatop);
deltop2=sigatop*lag2(sigatop);



delbot=sigatop;*/(P(k)-P(k-1)-muhat))^2;*/denominator used in estimation delta;
sigctop=((mpret3-3*muhat)**2);*/squared deviation from twice the mean return (for
ten-minute return);
run;

proc means data=mpret30min2 noprint;
by ric;
var sigatop sigctop deltop1 deltop2  delbot;
output out=varrat10min_30min
sum=sigatop sigctop deltop1 deltop2  delbot;
run;

data a.varrat10min_30min;
set varrat10min_30min;
by ric;
q = 3;  qm1 = q - 1;
        j = 1;
        theta=0;
        array deltaj(2) deltop1 -- deltop2;
        array delta(2);
        nq = _freq_;
        m = q*(nq-q+1)*(1-q/nq);
        siga = sigatop/(nq-1);
        sigc = sigctop/m;
        varrat10min_30min = sigc/siga;
        do until (j > qm1);
                delta(j) = nq*deltaj(j)/(delbot**2);
```

```
			theta = theta + ((2*(q-j)/q)**2)*delta(j);
			j+1;
		end;

   invVR10min_30min=abs(1-varrat10min_30min);
		z = sqrt(nq)*(varrat10min_30min-1)/sqrt(theta);


		keep ric nq varrat10min_30min invVR10min_30min z;
		label
		nq = "Number of Weekly Returns"
		varrat10min_30min = "VR(10,30)"
		z = "Heteroskedastic Robust Test Statistic"
   invVR10min_30min="/1-VR(10,30)/";
		run;

*proc print data=varrat10min_30min label noobs;
		*run;
*****************************************************

3.2.2 Variance Ratio(10min,60min)

*****************************************************;
		title2'Variance Ratio for 60-minute Returns to 10-minute returns';

data mpret60min2;
if _n_=1 then set muhat;
set mpret10min2;
by ric;
sigatop=((mpret-muhat)**2);*/(P(k)-P(k-1)-muhat))^2;*/squared deviations from mean
return;

*/sum((P(k)-P(k-1)-muhat)^2 * (P(k-j)-P(k-j-1)-muhat)^2);*/product of current and
lagged squared deviation;
deltop1=sigatop*lag(sigatop);
deltop2=sigatop*lag2(sigatop);
deltop3=sigatop*lag3(sigatop);
deltop4=sigatop*lag4(sigatop);
deltop5=sigatop*lag5(sigatop);



delbot=sigatop;*/(P(k)-P(k-1)-muhat))^2;*/denominator used in estimation delta;
sigctop=((mpret6-6*muhat)**2);*/squared deviation from twice the mean return (for
ten-minute return);
run;

proc means data=mpret60min2 noprint;
by ric;
var sigatop sigctop deltop1 deltop2 deltop3 deltop4 deltop5   delbot;
```

```
output out=varrat10min_60min
sum=sigatop sigctop deltop1 deltop2 deltop3 deltop4 deltop5   delbot;
run;

data a.varrat10min_60min;
set varrat10min_60min;
by ric;
q = 6;  qm1 = q - 1;
        j = 1;
        theta=0;
        array deltaj(5) deltop1 -- deltop5;
        array delta(5);
        nq = _freq_;
        m = q*(nq-q+1)*(1-q/nq);
        siga = sigatop/(nq-1);
        sigc = sigctop/m;
        varrat10min_60min = sigc/siga;
        do until (j > qm1);
                delta(j) = nq*deltaj(j)/(delbot**2);
                theta = theta + ((2*(q-j)/q)**2)*delta(j);
                j+1;
        end;

   invVR10min_60min=abs(1-varrat10min_60min);
        z = sqrt(nq)*(varrat10min_60min-1)/sqrt(theta);


        keep ric nq varrat10min_60min  invVR10min_60min z;
        label
        nq = "Number of Weekly Returns"
        varrat10min_60min = "VR(10,60)"
        z = "Heteroskedastic Robust Test Statistic"
   invVR10min_60min="/1-VR(10,60)/";
        run;

*proc print data=varrat10min_60min label noobs;
        *run;
data a.variance_ratio;
set a.varrat5min_30min a.varrat5min_60min a.varrat10min_30min
a.varrat10min_60min;
by ric;
drop nq;
run;

****************************************************************

4.0 Autocorrelation

****************************************************************
```

```
4.1construct 30-minute time interval;

*************************************************************;

title1'Return Autocorrelation';


data interval3;
        informat start time11.2 end time11.2;
        input start end;
        format start time11.2 end time11.2;
        datalines;
9:30:00        10:00:00
10:00:00 10:30:00
10:30:00 11:00:00
11:00:00 11:30:00
11:30:00 12:00:00
12:00:00 12:30:00
12:30:00 13:00:00
13:00:00 13:30:00
13:30:00 14:00:00
14:00:00 14:30:00
14:30:00 15:00:00
15:00:00 15:30:00
15:30:00 16:00:00
;
proc sql;
  create table ricdi3 as
  select *
  from date, interval3 ;
quit;

proc sort data=ricdi3;
by ric date start;
run;
* combine date & time;

data ricdti3;
        set ricdi3;
        dtstart = dhms(date,0,0,start);
        dtend = dhms(date,0,0,end);
        format dtstart datetime20.2 dtend datetime20.2;
        dt = dtend;
        keep ric date dtstart dtend dt;
        run;

proc sort data=ricdti3;
        by ric dtstart;
        run;
**construct thirty-minute time interval data;
```

```
data interval3;
	set ricdti3 data2;
	run;

proc sort data=interval3;
	by ric dt;
	run;

data a.interval30min (keep=ric date dtstart dtend endprice endmidpoint db ds nb ns vb
vs);
	set interval3;
	by ric;
	retain dtstart dtend db ds nb ns vb vs endprice endmidpoint;
	if _N_=1 then do; db=0; ds=0; nb=0; ns=0; vb=0; vs=0; end;

	if dtstart = . then
		do;
		if tradesign = 1 then
			do;
			db = db + price*volume;
			nb = nb + 1;
			vb = vb + volume;
			end;
		if tradesign = -1 then
			do;
			ds = ds + price*volume;
			ns = ns + 1;
			vs = vs + volume;
			end;
			endprice=price;
			endmidpoint=currentmidpoint;
		end;
	else
		do; output; db=0; ds=0; nb=0; ns=0; vb=0; vs=0; end;
	run;

*proc print data=a.interval30min (obs=100);*run;
**********************************************************

4.2 Compute midpoint return autocorrelation

**********************************************************;

proc sort data=a.interval30min;
by ric dtstart dtend;
run;
```

```
data mpret30min3;
set a.interval30min;
by ric date;
retain n;
  mpret=log(endmidpoint / lag(endmidpoint));*/logged thirty minutes return;
if first.ric or first.date then do;
   n=0;
   end;
   n+1;

if n<=1 then  mpret= .;

run;
*proc print data=mpret30min3 (obs=20);*run;


*4.2.1 30 minute midpoint return autocorrelation;

title2 '30 minute midpoint return autocorrelation';

data mpret30minautoreg;
set mpret30min3;
by ric;
mpretlag=lag(mpret);

run;

**30 minute midpoint return autocorrelation;


proc autoreg data=mpret30minautoreg;
model mpret = mpretlag /archtest dw=4 dwprob noint;
by ric;
ods output ParameterEstimates=a.ThirtyMinAutoreg ;
quit;
run;
data a.ThirtyMinAutoreg;
set a.ThirtyMinAutoreg;
*rename Estimate=AR_thirty;
abs_AR_thirty=abs(Estimate);
label abs_AR_thirty='/AR30/';
run;

**4.2.2 10 minute midpoint return autocorrelation;

title2 '10 minute midpoint return autogression';

data mpret10minautoreg;
set mpret10min2;
by ric;
```

```
mpretlag=lag(mpret);

run;

proc autoreg data=mpret10minautoreg;
model mpret = mpretlag /archtest dw=4 dwprob noint;
by ric;
ods output ParameterEstimates=a.TenMinAutoreg;
quit;
run;
data a.TenMinAutoreg;
set a.TenMinAutoreg;
*rename Estimate=AR_thirty;
abs_AR_ten=abs(Estimate);
label abs_AR_ten='/AR10/';
run;

**4.2.3 5 minute midpoint return autocorrelation;

title2 '5 minute midpoint return autogression';

data mpret5minautoreg;
set mpret5min;
by ric;
mpretlag=lag(mpret);

run;

proc autoreg data=mpret5minautoreg;
model mpret = mpretlag /archtest dw=4 dwprob noint;
by ric;
ods output ParameterEstimates=a.FiveMinAutoreg;
quit;
run;
data a.FiveMinAutoreg;
set a.FiveMinAutoreg;
*rename Estimate=AR_thirty;
abs_AR_Five=abs(Estimate);
label abs_AR_five='/AR5/';
run;


*****************************************************

5.0 adjPIN and PSOS (Duarte and Young (2009))

*****************************************************;

*5.1 Compute number of sells and number of buys each day
```

```
*****************************************************;

title 'Daily number of buy and sell';

data nbuy_nsell;
merge a.interval30min ricdi3(keep=ric date);
by ric;
format Date date9.;
run;

proc sort data=nbuy_nsell;
by ric date;
run;

proc means data=nbuy_nsell noprint;
by ric date;
var nb ns db ds vb vs;
output out=a.dailynbuy_nsell sum= nb ns db ds vb vs;
title2 'Daily number of buys and sells';
run;

data a.dailynbuy_nsell;
set a.dailynbuy_nsell;
periodm = year(date)*100+month(date);
buys=nb;sells=ns;
drop _type_ _freq_ nb ns db ds vb vs;
run;

proc sort data=a.dailynbuy_nsell;
by ric periodm;
run;

*Statistic description; *Correlation between buys and sells;

proc means data=a.dailynbuy_nsell mean var P95 P75 median P25 P5 ;
var buys sells;
output out=temp;
run;
proc corr data=a.dailynbuy_nsell;
var buys;
with sells;
run;



************************************************************

5.2 Compute adjPIN and PSOS by Duarte and Young (2009, JFE) Why is PIN Priced?

************************************************************;
```

```
proc printto print="G:\nyseJanMar2007 working file\data.out"; run;


ods output AdditionalEstimates=a.pin ConvergenceStatus=a.cs
 IterHistory=a.ih FitStatistics=a.fs;
proc nlmixed data=a.dailynbuy_nsell fd=central technique=quanew update=bfgs ;
 by ric ;
 parms za=-1.5 .5 1.5, zd=-1.5 .5 1.5, zub=1 4 7 , zus=1 4 7 , zeb=1 4 7 ,
                    zes=1 4 7 , zdelb=1 4 7 ,zdels=1 4 7 , zt=-1.5 .5 1.5 ;

 a = exp(za)/(1+exp(za));
 d = exp(zd)/(1+exp(zd));
 ub = exp(zub);
 us = exp(zus);
 eb = exp(zeb);
 es = exp(zes);
 delb = exp(zdelb);
 dels = exp(zdels);
 t = exp(zt)/(1+exp(zt));

 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 'ub' ub;
 estimate 'us' us;
 estimate 'eb' eb;
 estimate 'es' es;
 estimate 'delb' delb;
 estimate 'dels' dels;
 estimate 'theta' t;
 estimate 'adjPIN' adjpin;
 estimate 'PSOS' psos;
run;
```

```
proc printto; run;

***extract adjpin and psos;

data a.adjpin;
set a.pin;
by ric;
if label="adjPIN";
adjpin=Estimate;
tValue_adjpin=tValue;
*label adjpin="adjPIN";

keep ric adjpin tValue_adjpin;
run;



data a.psos;
set a.pin;
by ric;
if label="PSOS";
psos=Estimate;
tValue_psos=tValue;
*label psos="PSOS";
keep ric psos tValue_psos;
run;

***extract estimated parameters ;

data alpha;
set a.pin;
by ric;
if label="alpha";
alpha=Estimate;

keep ric alpha;
run;
data delta;
set a.pin;
by ric;
if label="delta";
delta=Estimate;

keep ric delta;
run;
data ub;
set a.pin;
by ric;
if label="ub";
```

```
ub=Estimate;

keep ric ub;
run;

data us;
set a.pin;
by ric;
if label="us";
us=Estimate;

keep ric us;
run;
data eb;
set a.pin;
by ric;
if label="eb";
eb=Estimate;

keep ric eb;
run;

data es;
set a.pin;
by ric;
if label="es";
es=Estimate;

keep ric es;
run;

data delb;
set a.pin;
by ric;
if label="delb";
delb=Estimate;

keep ric delb;
run;

data dels;
set a.pin;
by ric;
if label="dels";
dels=Estimate;

keep ric dels;
run;
data theta;
set a.pin;
```

```
by ric;
if label="theta";
theta=Estimate;

keep ric theta;
run;

data a.EstimatedParametersForadjPIN;
merge alpha delta ub us eb es delb dels theta a.adjpin a.psos;
by ric;
run;

*Statistic description about estimated parameters of adjPIN;

proc means data=a.EstimatedParametersForadjPIN P95 P75 median P25 P5;
var alpha delta ub us eb es delb dels theta adjpin tValue_adjpin psos  tValue_psos;
run;




*******************************************************************
***********************;

*Alternative way of computing adjPIN and PSOS by NLP;

*******************************************************************
************************;

*proc nlp data=a.dailynbuy_nsell fd=central technique=quanew update=bfgs noprint
out=a.adjpin_psos (keep=ric  a d ub us eb es delb dels t adjpin psos);
 *by ric ;

 *parms za=-1.5 .5 1.5, zd=-1.5 .5 1.5, zub=1 4 7 , zus=1 4 7 , zeb=1 4 7 ,
                    zes=1 4 7 , zdelb=1 4 7 ,zdels=1 4 7 , zt=-1.5 .5 1.5 ;
 *max loglik;

 *a = exp(za)/(1+exp(za));
 *d = exp(zd)/(1+exp(zd));
 *ub = exp(zub);
 *us = exp(zus);
 *eb = exp(zeb);
 *es = exp(zes);
 *delb = exp(zdelb);
 *dels = exp(zdels);
 *t = exp(zt)/(1+exp(zt));

 *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);


*run;

*****************************************************************************
*********************************;




*** Compute means of adjPIN and PSOS;

*proc means data=a.adjpin_psos noprint n mean std var;
*by ric;
*var adjpin psos;
*output out=a.muadjpin_and_mupsos (drop=_type_ _freq_) mean= mu_adjpin mu_psos;
*run;



*****************************************************************************

6.0 Pure Liquidity decomposition

*****************************************************************************;

proc sort data=a.spread;
by ric date;
run;

proc means data=a.spread mean n var std noprint;
by ric;
var effsprd;
output out=mu_effsprd  mean=mu_effsprd std=std_effsprd;
run;
proc means data=a.spread mean n var std noprint;
by ric;
var asprd;
output out=mu_asprd  mean=mu_asprd std=std_asprd;
```

```
run;
proc means data=a.spread mean n var std noprint;
by ric;
var rsprd;
output out=mu_rsprd  mean=mu_rsprd std=std_rsprd;
run;


data a.mu_spread ;
merge mu_effsprd mu_asprd mu_rsprd ;
by ric;
drop _type_ _freq_;


label mu_effsprd='average effective spread';
label std_effsprd='average standard deviation of effective spread';
label mu_asprd='average bid-ask spread';
label std_asprd='average standard deviation of bid-ask spread';
label mu_rsprd='average relative spread';
label std_rsprd='average standard deviation of relative spread';
run;

*data a.liqandinfor;
*merge a.muadjpin_and_mupsos a.mu_spread;
*by ric;

*mu_effsprd=mu_effsprd*100;
*std_effsprd=std_effsprd*100;
*mu_asprd=mu_asprd*100;
*std_asprd=std_asprd*100;
*mu_asprd=mu_rsprd*100;
*std_asprd=std_rsprd*100;
*mu_psos=mu_psos*100;
*mu_adjpin=mu_adjpin*100;
*keep ric  mu_effsprd std_effsprd mu_psos mu_adjpin mu_asprd std_asprd;
*run;


*proc reg data=a.liqandinfor outest=est;
*by ric;
*model mu_effsprd= mu_adjpin /noint ;
*ods output ParameterEstimates=REG_ParameterEstimates;
*quit;
*run;

*data temp;
*set REG_ParameterEstimates a.liqandinfor;
*Normalized_liq_error=stderr*100/std_effsprd;
*run;
```

```
****************************************************************************
********************

7.0 Relation Between  Pure Liquidity,  Asymmetric Information And Market
Efficiency,.

****************************************************************************
*********************;

data a.liq_infor_mkt;
merge a.adjpin a.psos a.normalized_pricing_error_STD  a.mu_spread
a.varrat5min_30min
a.varrat5min_60min a.varrat10min_30min a.varrat10min_60min a.tenminautoreg
a.thirtyminautoreg;
by ric;

**change variables into %;

adjpin=100*adjpin;
psos=100*psos;
mu_rsprd=100*mu_rsprd;
mu_asprd=100*mu_asprd;
mu_effsprd=100*mu_effsprd;
invVR5min_30min=100*invVR5min_30min;
invVR5min_60min=100*invVR5min_60min;
invVR10min_30min=100*invVR10min_30min;
invVR10min_60min=100*invVR10min_60min;
abs_AR_ten=100*abs_AR_ten;
abs_AR_thirty=100*abs_AR_thirty;


keep ric adjpin psos NormalisedPricingError mu_rsprd mu_asprd mu_effsprd
invVR5min_30min invVR5min_60min
invVR10min_30min invVR10min_60min abs_AR_ten abs_AR_thirty;
run;
proc sort data=a.liq_infor_mkt;
by ric;
run;

***export data to excel;
*proc export data=a.liq_infor_mkt
outfile="D:\nyse\liq_infor_nkt"
DBMS=excel replace;
*run;
proc corr data=a.liq_infor_mkt;
var invVR5min_30min invVR5min_60min
invVR10min_30min invVR10min_60min abs_AR_ten abs_AR_thirty
NormalisedPricingError mu_rsprd mu_asprd mu_effsprd adjpin psos  ;
quit;
run;
```

```
title1'Liquidity information asymmetry and market efficiency';



OPTIONS compress=yes ;
libname a 'D:\nyseJanMar2007 working file\results\main reg and corr';

***combines into one dataset;

data a.LiqInfoMkt_tot;
set a.liq_infor_mkt1 a.liq_infor_mkt2 a.liq_infor_mkt3 a.liq_infor_mkt4
a.liq_infor_mkt5 a.liq_infor_mkt6 a.liq_infor_mkt7 a.liq_infor_mkt8
a.liq_infor_mkt9 a.liq_infor_mkt10 a.liq_infor_mkt11;
if psos= . then delete;
if adjpin= . then delete;
run;
**Find firms with missing psos and adjpin;
data a.missing;
set a.liq_infor_mkt1 a.liq_infor_mkt2 a.liq_infor_mkt3 a.liq_infor_mkt4
a.liq_infor_mkt5 a.liq_infor_mkt6 a.liq_infor_mkt7 a.liq_infor_mkt8
a.liq_infor_mkt9 a.liq_infor_mkt10 a.liq_infor_mkt11;
if psos= . ;
if adjpin= . ;
run;
proc sort data=a.LiqInfoMkt_tot;
by ric;
run;

OPTIONS compress=yes ;
libname a 'D:\nyseJanMar2007 working file\results\buy_sell';
data a.dailynbuy_nsell_tot;
set a.dailynbuy_nsell1 a.dailynbuy_nsell2 a.dailynbuy_nsell3 a.dailynbuy_nsell4
a.dailynbuy_nsell5
a.dailynbuy_nsell6 a.dailynbuy_nsell7 a.dailynbuy_nsell8 a.dailynbuy_nsell9
a.dailynbuy_nsell10 a.dailynbuy_nsell11;
if ric='RNT.N' then delete;
if ric='PCU.N' then delete;
run;
proc sort data=a.dailynbuy_nsell_tot;
by ric;
run;

proc means data=a.dailynbuy_nsell_tot mean std var;
by ric;
var buys sells;
output out=a.Avgnbuy_nsell_tot mean=mu_buys mu_sells var=var_buys var_sells;
run;
```

```
proc corr data=a.dailynbuy_nsell_tot;
by ric;
var buys;
with sells;
ods output pearsonCorr=PearsonCorr;
run;
data a.buysell_corr;
set PearsonCorr;
BuySellCorre=buys;
P_value=pbuys;
keep ric BuySellCorre P_value;
run;

data a.summaryBuySell;
merge a.Avgnbuy_nsell_tot a.buysell_corr;
by ric;
run;

*proc corr data=a.Avgnbuy_nsell_tot;
*var mu_buys mu_sells;
*run;

proc export data=a.summaryBuySell
outfile="D:\nyseJanMar2007 working file\results\buy_sell\avebuysell"
DBMS=excel replace;
run;


OPTIONS compress=yes ;
libname a 'D:\nyseJanMar2007 working file\results\VR and test\VR(5,30)';
data a.VR5min30min_tot;
set a.varrat5min_30min1 a.varrat5min_30min2 a.varrat5min_30min3
a.varrat5min_30min4
a.varrat5min_30min5 a.varrat5min_30min6 a.varrat5min_30min7 a.varrat5min_30min8
a.varrat5min_30min9 a.varrat5min_30min10 a.varrat5min_30min11;
if ric='RNT.N' then delete;
if ric='PCU.N' then delete;
run;
proc sort data=a.VR5min30min_tot;
by ric;
run;
proc means data=a.VR5min30min_tot mean std var;
var invVR5min_30min z;
output out=a.VR5min30min_Test mean=muinvVR5min_30min mu_z;
run;

proc export data=a.VR5min30min_Test
outfile="D:\nyseJanMar2007 working file\results\VR and test\VR(5,30)\VR(5,30)test"
DBMS=excel replace;
run;
```

```
OPTIONS compress=yes ;
libname a 'D:\nyseJanMar2007 working file\results\VR and test\VR(5,60)';
data a.VR5min60min_tot;
set a.varrat5min_60min1 a.varrat5min_60min2 a.varrat5min_60min3
a.varrat5min_60min4
a.varrat5min_60min5 a.varrat5min_60min6 a.varrat5min_60min7 a.varrat5min_60min8
a.varrat5min_60min9 a.varrat5min_60min10 a.varrat5min_60min11;
if ric='RNT.N' then delete;
if ric='PCU.N' then delete;
run;
proc sort data=a.VR5min60min_tot;
by ric;
run;
proc means data=a.VR5min60min_tot mean std var;
var invVR5min_60min z;
output out=a.VR5min60min_Test mean=muinvVR5min_60min mu_z;
run;
proc export data=a.VR5min60min_Test
outfile="D:\nyseJanMar2007 working file\results\VR and test\VR(5,60)\VR(5,60)test"
DBMS=excel replace;
run;


OPTIONS compress=yes ;
libname a 'D:\nyseJanMar2007 working file\results\VR and test\VR(10,30)';
data a.VR10min30min_tot;
set a.varrat10min_30min1 a.varrat10min_30min2 a.varrat10min_30min3
a.varrat10min_30min4
a.varrat10min_30min5 a.varrat10min_30min6 a.varrat10min_30min7
a.varrat10min_30min8
a.varrat10min_30min9 a.varrat10min_30min10 a.varrat10min_30min11;
if ric='RNT.N' then delete;
if ric='PCU.N' then delete;
run;
proc sort data=a.VR10min30min_tot;
by ric;
run;
proc means data=a.VR10min30min_tot mean std var;
var invVR10min_30min z;
output out=a.VR10min30min_Test mean=muinvVR10min_30min mu_z;
run;
proc export data=a.VR10min30min_Test
outfile="D:\nyseJanMar2007 working file\results\VR and
test\VR(10,30)\VR(10,30)test"
DBMS=excel replace;
run;


OPTIONS compress=yes ;
libname a 'D:\nyseJanMar2007 working file\results\VR and test\VR(10,60)';
data a.VR10min60min_tot;
```

```
set a.varrat10min_60min1 a.varrat10min_60min2 a.varrat10min_60min3
a.varrat10min_60min4
a.varrat10min_60min5 a.varrat10min_60min6 a.varrat10min_60min7
a.varrat10min_60min8
a.varrat10min_60min9 a.varrat10min_60min10 a.varrat10min_60min11;
if ric='RNT.N' then delete;
if ric='PCU.N' then delete;
run;
proc sort data=a.VR10min60min_tot;
by ric;
run;
proc means data=a.VR10min60min_tot mean std var;
var invVR10min_60min z;
output out=a.VR10min60min_Test mean=muinvVR10min_60min mu_z;
run;
proc export data=a.VR10min60min_Test
outfile="D:\nyseJanMar2007 working file\results\VR and
test\VR(10,60)\VR(10,60)test"
DBMS=excel replace;
run;


OPTIONS compress=yes ;
libname a 'H:\nyseJanMar2007 working file\results\AR and test\AR5';
data a.AR5_tot;
set a.fiveminautoreg1 a.fiveminautoreg2 a.fiveminautoreg3 a.fiveminautoreg4
a.fiveminautoreg5
a.fiveminautoreg6 a.fiveminautoreg7 a.fiveminautoreg8 a.fiveminautoreg9
a.fiveminautoreg10 a.fiveminautoreg11;
if ric='RNT.N' then delete;
if ric='PCU.N' then delete;
run;
proc sort data=a.AR5_tot;
by ric;
run;
proc means data=a.AR5_tot mean std var;
var abs_AR_five tValue probt;
output out=a.AR5_Test mean=abs_AR5 mu_tValue probt;
run;
proc export data=a.AR5_Test
outfile="H:\nyseJanMar2007 working file\results\AR and test\AR5\AR5test"
DBMS=excel replace;
run;


OPTIONS compress=yes ;
libname a 'D:\nyseJanMar2007 working file\results\AR and test\AR10';
data a.AR10_tot;
set a.tenminautoreg1 a.tenminautoreg2 a.tenminautoreg3 a.tenminautoreg4
a.tenminautoreg5
```

```
a.tenminautoreg6 a.tenminautoreg7 a.tenminautoreg8 a.tenminautoreg9
a.tenminautoreg10 a.tenminautoreg11;
if ric='RNT.N' then delete;
if ric='PCU.N' then delete;
run;
proc sort data=a.AR10_tot;
by ric;
run;
proc means data=a.AR10_tot mean std var;
var abs_AR_ten tValue probt;
output out=a.AR10_Test mean=abs_AR10 mu_tValue probt;
run;
proc export data=a.AR10_Test
outfile="D:\nyseJanMar2007 working file\results\AR and test\AR10\AR10test"
DBMS=excel replace;
run;


OPTIONS compress=yes ;
libname a 'D:\nyseJanMar2007 working file\results\AR and test\AR30';
data a.AR30_tot;
set a.thirtyminautoreg1 a.thirtyminautoreg2 a.thirtyminautoreg3 a.thirtyminautoreg4
a.thirtyminautoreg5
a.thirtyminautoreg6 a.thirtyminautoreg7 a.thirtyminautoreg8 a.thirtyminautoreg9
a.thirtyminautoreg10 a.thirtyminautoreg11;
if ric='RNT.N' then delete;
if ric='PCU.N' then delete;
run;
proc sort data=a.AR30_tot;
by ric;
run;
proc means data=a.AR30_tot mean std var;
var abs_AR_thirty tValue probt;
output out=a.AR30_Test mean=abs_AR30 mu_tValue probt;
run;
proc export data=a.AR30_Test
outfile="D:\nyseJanMar2007 working file\results\AR and test\AR30\AR30test"
DBMS=excel replace;
run;
OPTIONS compress=yes ;
libname a 'D:\nyseJanMar2007 working file\results\Adjpin and PSOS';
data a.AdjpinAndPSOS_tot;
set a.estimatedparametersforadjpin1 a.estimatedparametersforadjpin2
a.estimatedparametersforadjpin3 a.estimatedparametersforadjpin4
a.estimatedparametersforadjpin5 a.estimatedparametersforadjpin6
a.estimatedparametersforadjpin7 a.estimatedparametersforadjpin8
a.estimatedparametersforadjpin9 a.estimatedparametersforadjpin10
a.estimatedparametersforadjpin11;
if ric='RNT.N' then delete;
if ric='PCU.N' then delete;
run;
```

```
proc sort data=a.AdjpinAndPSOS_tot;
by ric;
run;
proc means data=a.AdjpinAndPSOS_tot mean std var;
var alpha delta ub us eb es delb dels theta adjpin tValue_adjpin psos  tValue_psos;;
output out=a.AdjpinAndPSOS_test ;
run;
proc export data=a.AdjpinAndPSOS_test
outfile="D:\nyseJanMar2007 working file\results\Adjpin and
PSOS\AdjpinandPSOS_test"
DBMS=excel replace;
run;
```

References

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

Affleck-Graves, J., Callahan, C. M. & Chipalkatti, N. (2002). Earnings Predictability, Information Asymmetry, and Market Liquidity. *Journal of Accounting Research, 40*(3), 561-583.

Aktas, N., de Bodt, E. & van Oppens, H. (2008). Legal insider trading and market efficiency. *Journal of Banking & Finance, 32*(7), 1379-1392.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets, 5*(1), 31-56.

Amihud, Y. & Mendelson, H. (1980). Dealership Market. *Journal of Financial Economics, 8*(1), 31-53.

Amihud, Y. & Mendelson, H. (1986). Asset Pricing and the Bid-Ask Spread. *Journal of Financial Economics, 17*(2), 223-249.

Banz, R. W. (1981). The Relationship.Between Return and Market Value of Common Stocks. *Journal of Financial Economics, 9*(1), 3-18.

Bessembinder, H. (2003a). Issues in assessing trade execution costs. *Journal of Financial Markets, 6*(3), 233-257.

Bessembinder, H. (2003b). Trade Execution Costs and Market Quality after Decimalization. *Journal of Financial & Quantitative Analysis, 38*(4), 747-777.

Beveridge, S. & Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle". *Journal of Monetary Economics, 7*(2), 151-174.

Boehmer, E., Kelley, E. & Hall, M. (2009). Institutional investors and the informational efficiency of prices. *Review of Financial Studies*, working paper.

Brennan, M. J. & Subrahmanyama, A. (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics, 41*(3), 441-464.

Campbell, J. Y., Lo, A. W. & Mackinlay, A. C. (1997). *The Econometrics of Financial Markets* (2th ed.). New Jersey: Princeton University Press.

Chakravarty, S., Wood, R. A. & van Ness, R. A. (2004). Decimals And Liquidity: A Study Of The NYSE. *Journal of Financial Research, 27*(1), 75-94.

Chen, W.-P., Chung, H., Lee, C. & Liao, W.-L. (2007). Corporate Governance and Equity Liquidity: Analysis of S&P transparency and disclosure rankings. *Corporate Governance: An International Review, 15*(4), 644-660.

Chordia, T., Roll, R. & Subrahmanyam, A. (2000). Commonality in liquidity. *Journal of Financial Economics, 56*(1), 3-28.

Chordia, T., Roll, R. & Subrahmanyam, A. (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics, 65*(1), 111-130.

Chordia, T., Roll, R. & Subrahmanyam, A. (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics, 76*(2), 271-292.

Chordia, T., Roll, R. & Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics, 87*(2), 249-268.

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

Easley, D., Kiefer, N. M., O'Hara, M. & Paperman, J. B. (1996). Liquidity, Information, and Infrequently Traded Stocks. *Journal of Finance, 51*(4), 1405-1436.

Easley, D. & O'Hara, M. (1987). Price, Trade Size, and Information in Securities Markets. *Journal of Financial Economics*, *19*(1), 69-90.

Ellis, K., Michaely, R. & O'Hara, M. (2000). The Accuracy of Trade Classification Rules: Evidence from Nasdaq. *Journal of Financial & Quantitative Analysis, 35*(4), 529-551.

Fama, E. F. & French, K. R. (1988). Permanent and Temporary Components of Stock Prices. *Journal of Political Economy, 96*(2), 246-273.

Fama, E. F. & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance*, *47*(2), 427-465.

Fishman, M. J. & Hagerty, K. M. (1992). Insider trading and the efficiency of stock prices. *RAND Journal of Economics*, *23*(1), 106-122.

Furfine, C. H. (2003). Decimalization and market liquidity. *Economic Perspectives, 27*(4), 2-12.

Glosten, L. R. & Harris, L. E. (1988). Estimating the Components of the Bid/Ask Spread. *Journal of Financial Economics*, *21*(1), 123-142.

Glosten, L. R. & Milgrom, P. R. (1985). Bid, Ask and transaction Prices in a Specialist Market With Heterogeneously Informed Traders. *Journal of Financial Economics*, *14*(1), 71-100.

Harris, L. (2003). *Trading and exchanges: Market microstructure for practitioners*. US: Oxford University Press.

Harris, M. & Raviv, A. (1993). Differences of opinion make a horse race. *Review of Financial Studies, 6*(3), 473-506.

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

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

Hasbrouck, J. (1993). Assessing the quality of a security market: A new approach to transaction-cost measurement. *Review of Financial Studies, 6*(1).

He, H. & Wang, J. (1995). Differential informational and dynamic behavior of stock trading volume. *Review of Financial Studies*, *8*(4), 919-972.

Kandel, E. & Pearson, N. D. (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy, 103*(4), 831-872.

Keiber, K. L. (2006). Insider Trading Rules and Price Formation in Securities Markets: An Entropy Analysis of Strategic Trading. *International Journal of Theoretical & Applied Finance*, *9*(8), 1215-1243.

Keim, D. B. & Madhavan, A. (1996). The upstairs market for large-block transactions: Analysis and measurement of price effects. *Review of Financial Studies, 9*(1), 1-36.

Kyle, A. S. (1984). Market structure, information, futures markets, and price formation. *International agricultural trade: Advanced readings in price formation, market structure, and price instability*, 45-64.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica*, *53*(6), 1315-1335.

Lee, C. M. C. (1993). Market Integration and Price Execution for NYSE-Listed Securities. *The Journal of Finance, 48*(3), 1009-1038.

Lee, C. M. C. & Ready, M. J. (1991). Inferring Trade Direction from Intraday Data. *Journal of Finance*, *46*(2), 733-746.

Leland, H. E. (1992). Insider trading: Should it be prohibited? *Journal of Political Economy*, *100*(4), 859-887.

Lo, A. W. & MacKinlay, A. C. (1988). Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test. *The Review of Financial Studies, 1*(1), 41-66.

Madhavan, A., Richardson, M. & Roomans, M. (1997). Why do security prices change? A transaction-level analysis of NYSE stocks. *Review of Financial Studies, 10*(4), 1035-1064.

Poterba, J. M., Lawrence H. & Summers, J. M. (1988). MEAN REVERSION IN STOCK PRICES: Evidence and Implications. *Journal of Financial Economics, 22*(1), 27-59.

Pukthuanthong-Le, K. & Visaltanachoti, N. (2009). Commonality in liquidity: Evidence from the Stock Exchange of Thailand. *Pacific-Basin Finance Journal, 17*(1), 80-99.

Reinganum, M. R. (1990). Market microstructure and asset pricing: An empirical investigation of NYSE and NASDAQ securities. *Journal of Financial Economics*, *28*(1/2), 127-147.

Roll, R., Schwartz, E. & Subrahmanyam, A. (2007). Liquidity and the Law of One Price: The Case of the Futures-Cash Basis. *Journal of Finance*, *62*(5), 2201-2234.

Sarkar, A. & Schwartz, R. A. (2009). Market Sidedness: Insights into Motives for Trade Initiation. *Journal of Finance, 64*(1), 375-423.

Subrahmanyam, A. (1991). Risk aversion, market liquidity, and price efficiency. *Review of Financial Studies*, *4*(3), 417-441.

Subrahmanyam, A. (2009). The implications of liquidity and order flows for neoclassical finance. *Pacific-Basin Finance Journal, 17*(5), 527-532.