THREE ESSAYS ON MARKET MICROSTRUCTURE

BY

SUKWON KIM

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Management

August, 2009

Nashville, Tennessee

Approved:

Professor Hans R. Stoll

Professor Craig M. Lewis

Professor Ronald W. Masulis

Professor Robert E. Whaley

Professor Richard H. Willis

ACKNOWLEDGEMENTS

I cannot thank enough my advisor Dr. Hans R. Stoll, my committee chair. Without his great guidance, this work would not have been possible. I feel so fortunate to have such a role model scholar as my advisor, and I would like to express my deepest gratitude for his patronage. I also want to appreciate the support I received from my dissertation committee members – Dr. Craig M. Lewis, Dr. Ronald W. Masulis, Dr. Robert E. Whaley, and Dr. Richard H. Willis. Their inputs were essential to this dissertation. I want to give special thanks to Dr. Ronald W. Masulis, the chair of finance Ph.D. program. He always gave me great care and encouragements, which enabled me to clear the hurdles during my doctoral study.

I am grateful for kindness and care I received from other Owen professors. Especially, Dr. Clifford A. Ball and Dr. William G. Christie gave me numerous supports when I needed them the most. I thank financial support of the Vanderbilt Graduate School Fellowship and Financial Markets Research Center of Owen School. I am also thankful to Christoph Schenzler, who prepared valuable datasets that could be readily used by researchers in Owen.

Most of all, I need to mention all the love I received from my family while I was writing this work. My parents provided me with wonderful support and trust. My lovely wife Jinsun always gave me the deepest understanding and the biggest encouragement. My son Suwon brightened my mind like sunshine, and my forthcoming child, Surin, brought much happiness to me.

TABLE OF CONTENTS

CHAPTER I

PRICE DISCOVERY FROM PEERS: PEER LIQUIDITY AND OWN VOLATILITY

I. Introduction

According to French and Roll (1986), two sources of market volatility are information and the error in estimating the information's value ("pricing error"). Pricing error arises from the difference between the current stock price and the true stock price, where the true stock price reflects all available public information. Every announcement generates pricing error because there is no accurate way to convert qualitative information into quantitative price level. Information can be differently interpreted by traders. Trading is actually a process to resolve differences of opinion. For example, Harris and Raviv (1993) model how differences in opinion create trading volume. Without differences of opinion, there will be no trade at all, because stock price would instantly reach true value, and additional trading would only incur transaction cost. Pricing error has not been regarded as an important factor in investment decisions, because the resulting volatility seemed to be small and short lived.[1] However, recent

[1] The literature on differences of opinion is related to pricing error, because it studies a temporary deviation of stock price from true price. Miller (1977) shows stock price can be overvalued if there is short sales constraint. Hong and Stein (2003) and Boehme, Danielsen, and Sorescu (2006) argue differences of opinion can make stock prices to deviate from true value. Differences of opinion after a public information arrival is studied in Harris and Raviv (1993) and Kim and Verrecchia (1994). These papers do not exclusively deal with pricing error, but they show that differences of opinion can generate some particular patterns in stock price and volume. Pricing error is also related to price discovery models such as Kyle (1985) and Easley and O'Hara (1992). However, price discovery models focus on private information, and assume there are two types of investors – informed and liquidity traders. These settings are inappropriate to be applied to pricing error, because pricing error stems from public information.

findings show that pricing error is a significant portion of overall stock volatility (Evans and Lyons 2008), and pricing error can have substantial effect on traders in today's market environment.

While many institutions develop and use mechanical trading strategies, a higher pricing error means that a major input to the trading equation – stock price – includes significant amount of error. Let's say bad news arrives and stock price drops 12% instantly. A program trading department of an investment bank sells its entire holdings because their model require a 'stop loss', when the price drops more than 10%. After a few hours, investors determine the news cannot have such large effect for the stock and price recovers to 5% drop. The bank mistakenly sold their position at a lower price due to pricing error. Even long term traders are exposed to pricing error, because they need to choose when and how to place their orders. While we often use market closing price to value a stock, actual orders can be executed in different trading hours, and exposure to pricing error can depend on order execution strategy. For example, order placements at market open is subject to larger pricing error, as documented in Amihud and Mendelson (1987) and Stoll and Whaley (1990). Pricing error is also of interest to traders who want to read inside information from price movements. Volatility created by pricing error can make it difficult to interpret stock prices. Research into the determinants of pricing error would be valuable.

This paper examine whether pricing error in a stock can be affected by the trading activities of other stocks. If public information affects multiple stocks, investors can filter out pricing error by consulting the prices of other stocks.[2] The stocks that share the same

[2] Veldkamp (2006) shows investors prefer information that can be applied to multiple assets, and such behavior generates comovement of asset prices. Chan (1993) argues market makers consult multiple stock

set of public information can be called as ‘peer stocks’. Earnings news from GM, for example, can affect the stock prices of Ford or Chrysler. When the effect of information is not restricted to one stock, investors can reduce a stock’s pricing error by consulting the prices of peer stocks. Suppose there is bad news for automobile industry during non-trading hours. At the next day’s market opening, Ford stock goes down 10% while GM stock goes down 2%. Such differences can occur since no investor can make perfectly accurate assessment about the impact of a new piece of information for a stock’s price. Assuming information has a similar effect on Ford and GM, the true value of the bad news will be closer to the average of two price movements. Hence, two stocks will mean-revert to minus 6% at the following trade. While Ford and GM stocks record their opening prices, Chrysler stock did not trade yet.[3] Traders of Chrysler can learn that true value of the information is near minus 6%, and adjust their trading activities based on these priors. GM, and Ford stocks exhibit more volatility compared to Chrysler, because they traded earlier. Based on this intuition, I construct a model on pricing error and the effect of peer stocks.

The model has several implications. First, individual pricing error is a decreasing function of peer stock trading activity. Investors learn the value of information from peer stock prices and reduce their own stock’s pricing error.[4] This process is like using a

---

prices to reduce pricing error of their own stock. Pasquariello and Vega (2008) argue informed traders place strategic orders on other stocks to camouflage their trading in one stock. Accounting papers including Han and Wild (1990) and Clinch and Sinclair (1987) find that one firm’s earnings information has an effect on other stocks’ prices. In the international finance literature, studies such as Karolyi and Stulz (2003), show that the information transfers can also occur across markets in different nations.

[3] Lo and MacKinlay (1990) and Brennan, Jegadeesh, and Swaminathan (1993) argue stocks have different adjustment speed to market wide information. During 1997 ~ 2002 in NASDAQ, the stocks do not open simultaneously on 9:30 am. Most stocks have several minutes between market opening and the time of the first trade.

[4] Chan (1993) has a model that market makers consult other stock prices to lower the pricing error of their own stock price.

sample mean to estimate the value of the population mean, where estimation error (pricing error) decreases in the sample size (number of peer stock prices). Second, the liquidity of peer stocks is beneficial. If a stock has many liquid stocks as its peers, investors have more opportunity to learn, because there are a larger price data of peer stocks. A stock may have a smaller pricing error even without own trading. Third, the model implies that late trading stocks have less pricing error than fast trading stocks. This is because the traders of late trading stocks can learn from prices of fast trading stocks and reduce pricing error. As a result, the first reactions of fast trading stocks contain more pricing error than those of late trading stocks. A lower liquidity does not necessarily mean a higher pricing error. I test the model's implications using NASDAQ opening prices. I find that stocks that have more learning opportunity from other stocks have a lower volatility, a smaller pricing error, and a weaker tendency to mean-revert.

This paper contributes to the literature by identifying a relation between individual pricing error and the trading activities of multiple stocks. The result supplements a growing literature on the interaction of multiple trading activities such as Pasquariello and Vega (2008). It also sheds additional lights on the relation between individual liquidity and market liquidity. Acharya and Pedersen (2005) show that correlation between two liquidities can affect asset price. I provide a basis why such correlation should exist. Implications of the model can be applied to the selection of stock execution strategy or market design. For example, investors can use trading activities of other stocks to estimate pricing error in a stock price. They can incorporate this information when they choose their order execution strategy. A trading system can provide the information of multiple (and possibly international) stock prices to help decisions. In market design, the model implies that partial trading halt is better than circuit breaker, because investors can consult the prices of other stocks and fix their once ‘irrational’ pricing. The model can also give additional insights in known empirical price patterns such as relation between trading hour and volatility. Amihud and Mendelson (1987) and Stoll and Whaley (1990) document abnormally high volatility at market open compared to close. This puzzle is revisited by Amihud and Mendelson (1991), Forster and George (1996), Madhavan and Sofianos (1998), and Stoll (2000). This paper shows that the abnormal volatility at open is related to pricing error and learning effect.

The rest of paper is organized as follows: Section 2 proposes the model on pricing error after public information arrival. Section 3 tests the empirical hypotheses derived from the model. Section 4 concludes.

II. Model of Pricing Error

A. New Information and First Reaction

The model starts with an assumption that one cannot perfectly assess the true value of new information for a stock’s price.[5] If investors can accurately assess its true value instantly, there should be no trading, because trading a stock creates no profit but incurs transaction costs. Assume that when public information is released, investors cannot accurately convert new information into price changes, but they can only make an estimate of the true value of the information. The setting is similar to Kim and Verrecchia (1994) who model trading activity after an earnings announcement or Chan (1993) who

[5] French and Roll (1986) is one of the early papers to make this assumption and Harris and Raviv (1993) construct a model on trading volume based on the assumption.

model market makers consulting price data of other stocks. I write the estimated value of the new information event as:

$$Info^i_j = v + \eta^i_j \tag{1}$$

$Info^i_j$ is investor *i*'s estimated value of the information for stock *j, v* is the true value of information, and η is an independent, identically distributed error term with mean value 0 and variance σ^2. The term η can be also thought of as a measure of investor differences of opinion. Note that *v* does not have a subscript *j*, because the value of information is assumed to be common to multiple stocks. I define peer stocks as the stocks that share the same information.

Investors know their estimates *Info* includes error η, but they cannot observe *v* or η. This setting assumes investors' ability to process information varies, but each investor cannot measure the size of her own estimation error. Each investor believes her estimate *Info* is an unbiased estimate of the true value *v*, so probability that the true value of the information is higher than her own estimate is 50%. If investors wait until the true value is revealed, their short term profit by trading the stock is negative, because they have to buy or sell at the correct price and pay a transaction cost. On the other hand, if they trade before the true value is revealed, there is a chance to get a better price than what they expected. For example, if an investor buys a stock sufficiently below her *Info*, her expected profit based on her own belief is: (*Info – transaction price – transaction cost* ≥ *0*). Hence, less risk-averse investors trade before the true value is revealed, and the number of those investors is proportional to the total number of investors in a stock. The size of the investor pool for a stock is given, and it varies across stocks. If one of those risk taking investors wants to buy a stock, her bid quote will be her estimate *Info* minus transaction cost *m* that must be incurred:

$$Bid_j^{buyer} = Info_j^{buyer} - m_j^{buyer} = v + \eta_j^{buyer} - m_j^{buyer} \tag{2}$$

Similarly, her ask quote will be:

$$Ask_j^{seller} = Info_j^{seller} + m_j^{seller} = v + \eta_j^{seller} + m_j^{seller}$$

These quotes represent the maximum or minimum price an investor is willing to pay based on her own belief. Because η is a distribution, there can be a case when a buyer's bid quote is higher than a seller's ask quote. Then a trade occurs and price change is approximately $v + (\eta_j^{buyer} + \eta_j^{seller})/2$, provided that the size of *m* is similar across investors who are trading the same stock. This price change contains error, because the estimation error η of the buyer and seller are not completely offset. We can rewrite the price change as the sum of the information's value and an error term *e*, which is a linear function of η:

$$\Delta P_i = v + e_j \tag{3}$$

Because η is independent and identically distributed with zero mean, *e* is also independent, identically distributed with zero mean. The stock price changes have a common true value *v* and an individual error term *e* as illustrated in figure 1.

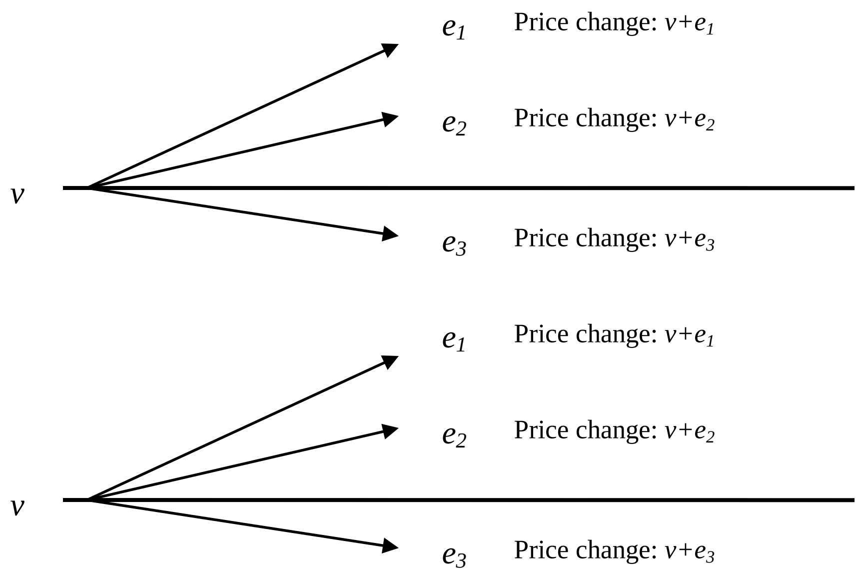


Figure 1. Price change after industry wide information arrival (Multiple stock view).

B. Price Discovery from Peers

If one investor knows the true information value *v*, she can buy undervalued stock and sell overvalued stock to earn arbitrage profit. By this arbitrage trading, the stock price change will eventually arrive at its true level *v*. This process is illustrated in figure 2.

In the real world, it is impossible to know the true value *v*. The second best method for investors is to estimate the true value *v* from the price changes of its peer stocks. The cross referencing is possible because all stocks in the peer group are affected by the same information set. In statistical sense, it is like using the sample mean in place of true mean. Figure 3 shows the case when traders use the average price changes of two peer stocks.

A strategy of selling the stock that moved higher than the peer stock average and buying the stock that moved lower than the average yields profits. This strategy yields a dynamic arbitrage, since a profit is guaranteed by the Law of Larger Numbers; the sample average converges to the population mean after a large number of trials. In this case, sample average corresponds to the average price changes of peers stocks, and population mean corresponds to the true value of an information event.

Result 1: A stock's price change after an information event's arrival converges to the average price change of its peer stocks.

Proof: See Appendix

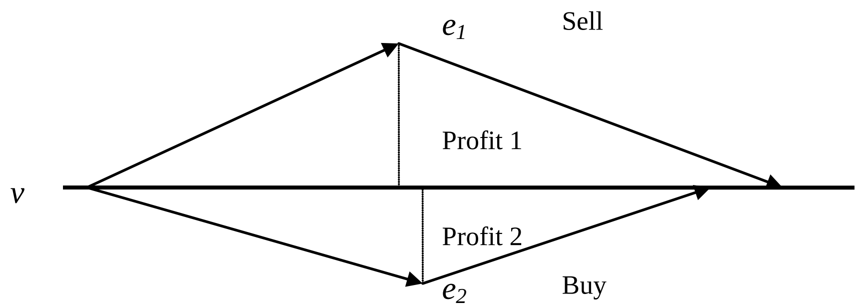


Figure 2. Arbitrage strategy when true value *v* is known.

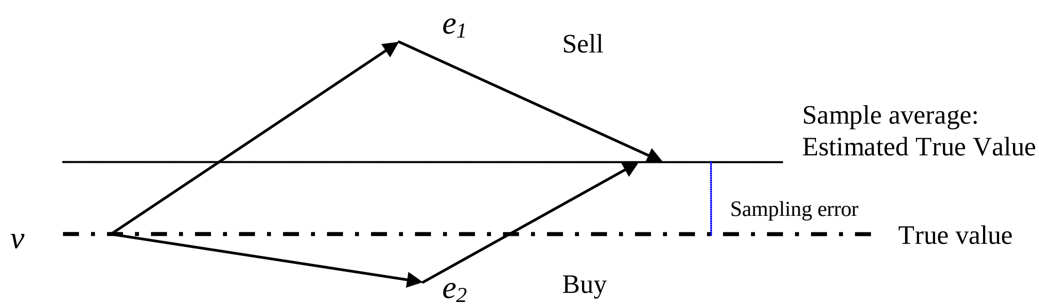


Figure 3. Estimating true value from price changes.

The average of price change is closer to the true value *v*, but it still contains sampling error. Investors can update their beliefs after observing the average price, but due to the sampling error, there are still differences of opinion about the true value of information. So some investors still trade based on their updated beliefs. If there are many peer stocks with trading, investors can extract the true value with higher accuracy, because the standard error of the estimate is a decreasing function of sample size.[6] This process continues until sufficient amount of trade data is accumulated and differences of opinion become small.

Result 2: A stock's price change after the information arrival becomes more accurate, the larger is the sample of previously traded peer stocks.

Proof: See Appendix

Result 2 shows that other things equal, a stock that does not have previous prices to consult will have larger error in its price changes. This result explains why market opening prices should have larger pricing error compared to closing prices. (Stoll (2000) calls this phenomenon as 'opening friction'.) Compare market opening price to the prices of other trading hours. At other trading hours, investors have continuous price change information of peer stocks as well as its own stock. At the market open, however, all stocks have not traded for hours.[7] Investors are forced to estimate the value of overnight

[6] Note that the cross consulting cannot reduce any bias in the value of information. If all the peer stocks contain the same amount of bias in the value *v*, comparing with other prices would still yield a biased result. Learning from peers can even create a bubble, by replicating and confirming the bias in peer stocks.
[7] Exceptions are stocks cross listed on exchanges that trades earlier.

information without the help of current peer stock prices. Opening price, therefore, has a higher variance compared to its prices at other hours of the trading day. The model not only explains the cause of the friction, but also predicts that stocks opening later than their peers would exhibit lower opening friction.

The learning framework implies that a trading halt of a few stocks is more beneficial than a circuit breaker. A trading halt gives investors an opportunity to compare a stock's price with the prices of other related stocks, and this can reduce its pricing error. The circuit breaker, on the other hand, does not reduce pricing error, because all the stocks in the market stop having price information.

C. Liquidity, Peer Stocks, and Volatility

Stocks do not always react to new information at the same time in actual trades. An example is opening prices. Most stocks open later than the official market opening time of 9:30 am. Stoll and Whaley (1990) document the average time elapse between the official opening of the exchange and the opening transaction in a stock was 15 minutes in 1986 for NYSE stocks. Data from 1997 to 2002 shows the average opening delay is about 6 minutes for NASDAQ stocks.[8]

Assume investors arrive sequentially to the market after a public announcement. Every short period of time s, one investor arrives at the market. s is decreasing in the size of the investor pool in a stock, so a stock with a larger investor pool has a higher arrival rate. An investor has one bid quote and one ask quote based on her belief. Her belief *Info* is based on a random draw from the distribution of η. The variance σ^2 is similar across

[8] This number excludes outliers that open later than 30 minutes.

stocks. The probability of a transaction after time t is the probability that the highest bid quote is above the lowest ask quote:

$$\Pr(\max(Bid) \geq \min(Ask)) = \Pr(\max(v + \eta_j^{buyer} - m_j^{buyer}) \geq \min(v + \eta_j^{seller} + m_j^{seller}))$$
$$= \Pr(\max(\eta_j^{buyer} - m_j^{buyer}) - \min(\eta_j^{seller} + m_j^{seller}) \geq 0) \qquad (4)$$

This probability is increasing in the number of quotes up to time t and the size of the estimation error σ^2, while decreasing in the size of the transaction cost m. The number of quotes is decreasing in s, because a stock with a slower arrival rate will have fewer investors willing to trade in the stock. Define a liquid stock as a stock with a larger investor pool and lower transaction cost. Then a liquid stock has a higher probability of a transaction during a fixed amount of time t. If peer stocks are relatively more liquid, investors would have a larger price data to update their estimations. Other things equal, a stock's pricing error is decreasing in the liquidity of peer stocks.

To illustrate the point, consider three peer stocks, A, B and C. Stock A and B are traded more frequently than stock C (higher liquidity). New information hits the market and stock A and B have immediate transactions, because they have more liquidity. Now stock C trades a minute later. Traders of stock C can learn from the prior price changes of stocks A and B. Hence, stock C's first price change after the information arrival can be more accurate, even if it did not have trading activity to resolve differences in opinion.

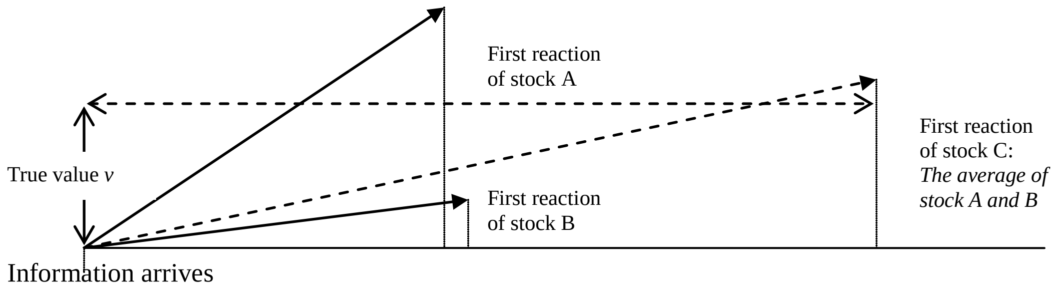


Figure 4. Price discovery process.

This analysis predicts price leadership by the more liquid stocks. When investors use peer stock prices to update their estimates, the average of the liquid stock transactions is an unbiased estimator of the true value. Hence, the average reaction of liquid stocks will lead the average reaction of illiquid stocks. Chordia and Swaminathan (2000) and Gervais, Kaniel and Mingelgrin (2001) empirically find such price leadership by liquid stocks. Note however, that the evidence of price leadership does not indicate that faster movers have always smaller pricing errors.

The price discovery model gives testable predictions from the learning effect. From *result 1*, stock price changes should show mean-reversion to peer stocks' average price change after new information arrival. From *result 2*, the error in the first reaction should be decreasing by the number of peer stocks' trades. Also, one can infer from the results that the degree of mean-reversion is weaker when a stock has more learning opportunity.

D. Scope of Peer Stocks

Result 1 predicts that stocks will mean-revert to the cross sectional average. Note that this result is based on the unrealistic assumption that all peer stocks react to an information event in the same direction. If firms are highly competitive, for example, good news for one firm can be bad news for the other firms. I relax this assumption and discuss the scope of peer stocks in this section.

Investors know a firm's characteristics before trading its stock. (Leverage, cost structure, or industry organization, etc.) I introduce a variable c, which represents the sensitivity of a stock to information, based on known firm characteristics. The sensitivity c can be positive or negative. For simplicity, let the effect of information and pricing error be proportional to the characteristic factor c. Then the price change of stock i is expressed as: $(\Delta P_i | c = c_i) = c_i \cdot (v + e_i)$. Similarly, the price change of a peer stock j is expressed as: $(\Delta P_j | c = c_j) = c_j \cdot (v + e_j)$. Investors can take the sensitivity out and directly compare $v+e_i$ to $v+e_j$, assuming they already know c_i and c_j. Investors of stock j can compare the price change of stock j to the price change of stock i by multiplying c_j/c_i.

$$(\Delta P_i | c = c_i) = c_i \cdot (v + e_i)$$

$$(\Delta P_j | c = c_j) = c_j \cdot (v + e_j)$$

$$(\Delta P_j | c = c_i) = [c_j \cdot (v + e_j)] \times {}^{c_i}\!/_{c_j} = c_i \cdot (v + e_j) \qquad (5)$$

Now the investors can follow the procedures in the previous section to reduce the pricing error. The basic idea is that investors account for the cross sectional difference in sensitivities to information. For example, investors of airline industry stocks can be surprised at a sudden increase in crude oil price. Observing the stock prices of refining companies can give investors some idea of whether the price change is temporary or not. Investors do not compare raw returns, because the sensitivity of the airline stock prices to crude oil price differs from that of refining firms stock prices. In this case, c is the dependence of the airline industry and the refining industry on crude oil price.

Any stock with pre-known characteristics can be useful for reducing the pricing error of other related stocks. Investors can use a large set of trade data to get better pricing of their own stock. Such trade data can include stock prices of major supplier industries, prices of derivatives, and prices of stocks on other exchanges. The range of peer stocks should not be restricted to close competitors or firms in the same industry. A policy implication is that it is better to provide more price information to investors. In my empirical work, I use equal-weighted market average return as a benchmark of the peer stock movements.[9] This setting would tend to capture the effect of market wide public information, which will in general have a similar effect across stocks. In other words, the effect of different characteristic *c* will be somewhat neutralized at the market wide level.

In the following section, I empirically test the results of the model using market opening prices. The market opening transaction is the first reaction to an overnight information arrival, and so I can test the model without the difficulty of identifying the information arrival time and the first reaction to it. I also test whether the model actually explains opening friction, documented by Amihud and Mendelson (1987), Stoll and Whaley (1990), and Stoll (2000). I analyze returns instead of price changes to compare with the earlier literature of opening friction. However, implications of the model are unchanged by using stock returns.[10]

III. Empirical Tests

A. Data

The main data source of this study comes from the Financial Markets Research Center (FMRC) in Owen Graduate School of Management, Vanderbilt University. FMRC has daily market microstructure database that is constructed from Trade and Quote (TAQ) data. The data covers all firms in TAQ except the stocks with daily prices

[9] Since the model gives the same weight for all prices, I use equal weighted return.
[10] Switching between two measures is not rare. For example, Chordia and Subrahmanyam (2004) models order imbalance using price changes, but use returns in empirical analysis.

below $3. Information in the database includes market microstructure variables such as time of the opening trade, bid-ask spread, dollar volume and price. I use daily data from January 1997 to July 2002 in this study, and from this point, I call this dataset the market microstructure (MMS) dataset. MMS dataset provides opening price, closing price and noon price. Opening price is the first traded price after official market opening (9:30 am). Closing price is the last traded price before official market closing. (4:00 pm) Noon price is the traded price closest to 12:00 pm. If a stock does not have more than 10 trades during a day, I drop that day's observation. This filter makes sure I use the stocks with considerable trading activity, and reduces the problem of infrequent trading. Additionally, I control for stock splits and dividends by deleting the returns in the window [-1, +1] of the event date. I call an individual stock's monthly variance based on open to open return as 'opening variance'. The closing variance is based on close to close return.

I use NASDAQ listed firms throughout the analysis.[11] NASDAQ data fits the model's learning framework well for several reasons. The model is based on a continuous trading framework, and for my sample period, NASDAQ has a continuous trading process at opening. In contrast, NYSE has a call auction at opening. The second reason is related to diversification. NASDAQ has relatively homogeneous firms compared to the NYSE. Many firms are in high-tech industries that share the same information set, so it is easier to learn from peer stocks. The third reason is the short speed it takes to reflect overnight information. It will be hard to observe the learning effect if it happens over long periods of time. Masulis and Shivakumar (2002) show NASDAQ stocks reflect overnight seasoned equity offering announcement an hour faster than NYSE stocks.

[11] MMS follows the TAQ's categorization for major exchange. MMS uses all the transactions from the listed exchanges to define opening, closing and noon price.

B. Mean-reversion

Since I analyze various stock returns around market opening, I define those stock returns first. Let the opening return be the return between two consecutive opening prices. I define the closing return be the return between two consecutive closing prices. The overnight return is the return between last day's closing price and today's opening price. The morning return is the return between today's opening price and today's noon price. The following figure shows the types of returns I use.

The overnight return corresponds to the first price change to new information arrival in the model, because the overnight return is the first price movement after overnight information arrivals. *Result 1* predicts that pricing errors should be reduced by using the movements of peer stocks. The simplest form of error correction process is mean-reversion to the market average movement.[12] I check the morning return to see how much of the overnight return is reverted to the market average movement. I convert *result 1* to *hypothesis 1* as follows:

Hypothesis 1: A stock's overnight return has a tendency to converge to the average return of the stocks at the following morning.

[12] This statement assumes the sign of sensitivity *c* is similar. Market wide information would generate a similar sign of *c* across many stocks.

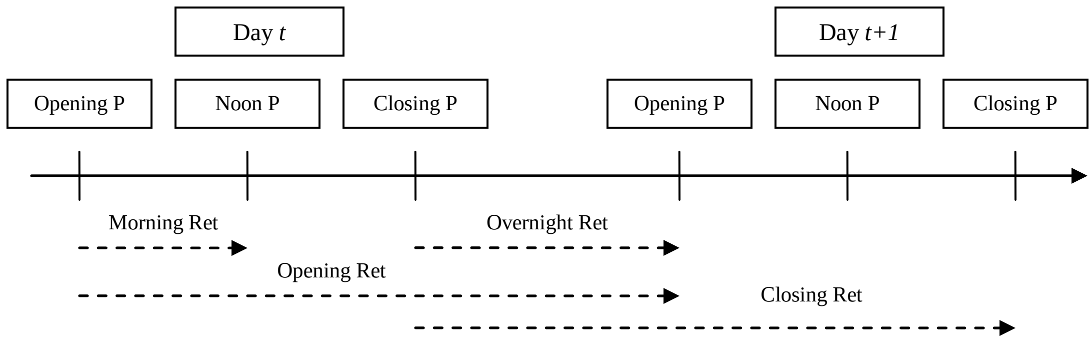


Figure 5. Types of returns.

The equation used to test the hypothesis 1 is as follows:

$$r^i_{morning} = \alpha + \lambda \cdot (r^i_{overnight} - \bar{r}_{overnight}) + \gamma_1 \cdot \bar{r}_{morning} \qquad (6)$$

- $r^i_{morning}$ is the return between today's first price and noon price
- $\bar{r}_{morning}$ is the average of morning returns
- $r^i_{overnight}$ is the return between previous day's closing price and today's opening price
- $\bar{r}_{overnight}$ is the average of overnight returns

To calculate the average overnight return for each day, I use the stocks that had first transaction in 1 minute after market open. These prices are the first reactions to overnight information without much learning effect. Under *Hypothesis 1*, if a stock has higher overnight return than average, its morning return should be lower. This prediction means the sign of coefficient λ should be negative. I estimate equation (6) for each stock, using all the time-series observations available, except cases when an opening price has a time stamp later than 10:00 am. Then I count the stocks with negative λs.

Table 1 shows the stocks in general have negative λs. A simple sign test confirms the significant tendency of λs to be negative. There is cross-sectional mean-reversion to the peer stocks' average. This result indicates there is considerable pricing error at market opening, and the error is reduced by converging to cross-sectional average. Substituting market-wide average of fast opening stocks with industry-wide average does not change the pattern.

Table 1

Mean-reversion

I run following regression for each stock every year and see the sign of lambda.

Model: $r^i_{morning} = \alpha + \lambda \cdot (r^i_{overnight} - \bar{r}_{overnight}) + \gamma_1 \cdot \bar{r}_{morning}$ (6)

- $r^i_{morning}$ is the return between today's first price and noon price
- $\bar{r}_{morning}$ is the average of morning returns
- $r^i_{overnight}$ is the return between previous day's closing price and today's opening price
- $\bar{r}_{overnight}$ is the average of overnight returns

The average overnight return is calculated from the overnight returns of the stocks that opened in the first minute after market open. Stocks are classified by average market value. Percent values are in the parentheses. Sign test shows the probability to have one type of sign occurring over 60% by chance is below 1%. Panel A shows the sum of all negative and positive coefficients, while panel B only uses coefficients significant in 5% level.

Panel A: Number of negative and positive lambda signs

	All stocks	Size (Min)	Size (2)	Size (3)	Size (Max)
Negative signs	1308	335	335	333	305
	(94.8%)	(97.1%)	(97.1%)	(96.5%)	(88.4%)
Positive signs	72	10	10	12	40
	(5.2%)	(2.9%)	(2.9%)	(3.5%)	(11.6%)
Total	1380	345	345	345	345
	(100%)	(100%)	(100%)	(100%)	(100%)

Panel B: Number of significant lambda signs in 5% level

	All stocks	Size (Min)	Size (2)	Size (3)	Size (Max)
Negative signs	1115	299	297	285	234
	(80.8%)	(86.7%)	(86.1%)	(82.6%)	(67.8%)
Positive signs	21	4	4	1	12
	(1.5%)	(1.1%)	(1.2%)	(0.3%)	(3.5%)
Insignificant	244	42	44	59	99
	(17.7%)	(12.2%)	(12.7%)	(17.1%)	(28.7%)
Total	1380	345	345	345	345
	(100%)	(100%)	(100%)	(100%)	(100%)

A limitation of the previous test is that it only counts the *frequency* of mean-reversion. In order to estimate the *degree* of mean-reversion, we need to compare how much of the overnight return is offset by the mean-reversion process. The degree of mean-reversion can be measured by the profit of an arbitrage trading strategy. The model predicts that investors can seek true value of information by buying the stocks that moved above average and selling the stocks that moved below average. Since it measures the differences between individual returns and average returns, the amount of profit acquired from the trading strategy will be similar to the amount of mean-reversion.

To make the trading strategy feasible, the average overnight return of a day is calculated from the stocks that had opening transactions in the first minute after market open. Then, I assume an arbitrager starts buying or selling the stocks that had first transactions two minutes after market open.[13] This setting gives the arbitrager some time to digest and use the prior price data. According to the model, if a stock's overnight return is above that day's average overnight return, the stock is likely to have positive pricing error. An arbitrager sells those stocks and buys the ones that moved below average. She puts equal weight in two portfolios, and the stocks in each portfolio also have equal weights. The equal amount of buying and selling implies a zero investment strategy. The profit is measured by comparing the two portfolio's morning return, which is the price change between opening price and noon price. I take the yearly average of the daily returns from this trading strategy.

Table 2 presents the result of the trading. For the all years in my dataset, the arbitrage trading earns profits in the range of 0.4 ~ 1.3%, which can exceed transaction costs. The profit gets lower as we go to more recent years, indicating the pricing error is decreasing

[13] The opening delay of a stock should be less than 30 minutes to be included in this analysis.

over time. Perhaps increased price information from other markets (such as international markets) and lower transaction cost for the arbitrage strategy have contributed to this phenomenon. The existence of arbitrage profits confirms that stocks overshoot or underreact to overnight information, and the pricing error is corrected by a subsequent mean-reversion process.

The model predicts that pricing error will be small for slow moving stocks, so we can infer that the profit of the arbitrage trading (degree of mean-reversion) would be small for a stock that reacts later to an information event. I test whether this is the case for opening prices. I use a stock's opening delay, which is the difference between the time of the first transaction and the official market opening, to measure the speed of a stock's reaction to overnight information. Each day I rank the opening delay into quartiles, and calculate the average profit of the arbitrage trading by the quartiles. Panel A of table 3 shows the result. Although the relation is not completely monotonic, we can see the profit is the lowest in the latest opening quartile, indicating the stocks with the largest learning opportunity have the least degree of mean-reversion. Panel B of Table 3 classifies stocks first by market value quartiles and then the opening delay quartiles. This process gives 4x4 = 16 clusters, and I calculate the profit of the arbitrage strategy by each cluster. Panel B of Table 3 shows that the profit of the arbitrage is decreasing in learning opportunity. The stocks with longer opening delay yields lower return in general. This result shows that the stocks with more learning opportunity have lower degree of mean-reversion.

Table 2

An arbitrage trading strategy using mean-reversion

I classify stocks into two categories using overnight returns. If overnight return is above that day's average overnight return, I assume an investor sells those stocks. The investor buys the ones that moved below average. The average is calculated from the overnight returns of the stocks that opened in the first minute after market open. The investor trades the stocks that opened later than 2 minutes after market open. The investor puts equal weight to two portfolios, and the stocks in each portfolio also have equal weights. The equal amount of buying and selling gives a zero investment strategy. The profit is measured by comparing two portfolio's morning return, which is the price change between opening price and noon price. I take yearly average of the daily return from the strategy in panel A. In Panel B, I take average of the daily return by quartiles of the time between market open and the first transaction.

Panel A: Arbitrage return by year

Year	Above average portfolio return "Sell" (A)	Below average portfolio return "Buy" (B)	Trading Profit Buy – Sell (B – A)
1997	-0.76%	0.53%	1.29%
1998	-0.34%	0.74%	1.08%
1999	-0.24%	0.59%	0.83%
2000	-0.18%	0.61%	0.79%
2001	-0.13%	0.25%	0.38%
2002	-0.19%	0.20%	0.40%

Table 3

Degree of mean-reversion and opening delay

In panel A, I rank the opening delay of stocks into quartiles and report the profit of the arbitrage trading. In panel B I divide stocks by market value quartiles, and then each market value quartile is divided into opening delay quartiles. This process gives 4x4 = 16 clusters, and I measure the profit of the arbitrage strategy in each cluster.

Panel A: Arbitrage return by opening delay

Quartiles of opening delay	Above average portfolio return "Sell" (A)	Below average portfolio return "Buy" (B)	Trading Profit Buy - Sell (B – A)
1 (Fastest)	-0.78%	0.31%	1.09%
2	-0.97%	0.48%	1.45%
3	-0.84%	0.36%	1.20%
4 (Slowest)	-0.36%	0.21%	0.57%

Panel B: Arbitrage return by market value and opening delay				
Quartiles of market value	Quartiles of opening delay	Above average portfolio return “Sell” (A)	Below average portfolio return “Buy” (B)	Trading Profit Buy - Sell (B – A)
1 (Smallest)	1 (Fastest)	-1.47%	0.77%	2.25%
1 (Smallest)	2	-1.36%	0.74%	2.10%
1 (Smallest)	3	-1.02%	0.50%	1.52%
1 (Smallest)	4 (Slowest)	-0.37%	0.20%	0.57%
2	1 (Fastest)	-1.09%	0.41%	1.50%
2	2	-1.15%	0.57%	1.72%
2	3	-0.90%	0.36%	1. 26%
2	4 (Slowest)	-0.37%	0.20%	0.57%
3	1 (Fastest)	-0.93%	0.32%	1.25%
3	2	-0.95%	0.44%	1.39%
3	3	-0.73%	0.30%	1.03%
3	4 (Slowest)	-0.37%	0.23%	0.60%
4 (Largest)	1 (Fastest)	-0.49%	0.19%	0.68%
4 (Largest)	2	-0.59%	0.33%	0.92%
4 (Largest)	3	-0.55%	0.27%	0.82%
4 (Largest)	4 (Slowest)	-0.31%	0.29%	0.60%

C. Learning Opportunity and Opening Return Volatility

The model predicts that learning effect reduces pricing error of a stock without trading the stock. For each stock, I count the number of prior opening transactions of other stocks every day. I call this variable '*prior-openings*'. If a stock opened later than 5 other stocks in a market, it would have 5 as prior-openings. The setting reflects the feature of the model that pricing error is decreasing in the number of prior peer transactions. Also note that, as in table 4, prior-openings variable is significantly negatively correlated with measures of liquidity.

To capture the pricing error at opening, I use the return volatility derived from the opening prices. Typically, the daily return is measured by the price change between two consecutive closing prices. Here two consecutive opening prices are used to calculate the daily return, which is then used to calculate monthly standard deviation of the return. The model predicts that the pricing error should be decreasing as investors have greater opportunity to learn from peer stock prices. The second hypothesis is formally stated as follows:

Hypothesis 2: Opening return volatility decreases in prior-openings of peer stocks.

As a dependent variable, I rank opening return variance into 11 categories and adjust the rank as follows:

$$Open\_vol = (Opening\_return\_categories/10) - 0.5$$

Table 4

Correlation among different liquidity variables

Prior-openings count the number of other stocks opened before a stock has its first trade. Bid-ask spread is the difference between quoted bid and ask prices. Dollar volume is calculated by a stock's closing price and volume. Depth is the sum of average bid size and ask size in share numbers. The probability that correlation coefficient is actually zero is in parentheses.

	Prior-openings	Bid-ask Spread	Dollar Volume	Depth
Prior-openings		0.28	-0.29	-0.21
		(0.00)	(0.00)	(0.00)
Bid-ask Spread	0.28			
	(0.00)			
Dollar Volume	-0.29			
	(0.00)			
Depth	-0.21			
	(0.00)			

This process is used in Mendenhall (2004) or Bernard and Thomas (1990) to control the skewness and outliers in the variable. The opening return volatility has range between -0.5 and 0.5. 0.1 is the difference between two consecutive deciles.

While the prior-openings variable is a daily value, the opening return volatility is a monthly statistic. The monthly average values of the prior-openings are used as an explanatory variable. Using monthly measures reduces the sample size and may reduce overall explanatory power of the model. The regression equation has a monthly panel data structure.

$$Open\_vol_{im} = \pi \cdot p_{im} + \Gamma \cdot Z_{im} + \varepsilon_{im} \qquad (7)$$

where p_{im} is average number of prior-openings and Z_{im} is a matrix of control variables

If the opening friction decreases in investors' learning opportunity, the sign of π should be negative. The control variables are: natural log of the market value, the opening volume (the number of shares traded in the first trade), the Herfindahl index of industry sales and a trading halt dummy. The log of market value is added to control for firm size and firm's sensitivity to overnight information. The market value is calculated using daily closing price and daily shares outstanding in the CRSP database. Opening volume controls the size of trading activity at the open. To account for the effect of industry structure, I include the Herfindahl index of sales, taken from Compustat quarterly database. The trading halt dummy takes value 1 if there is any trading halt for the stock during the month. Trading halts usually occur when there is special information event in a stock, so the stock experiencing a halt may have extra volatility. Except for the Herfindahl index and trading halt dummy, the other control variables are daily values, so I use the monthly averages of them in the regressions.

To control cross-sectional heteroskedasticity and serial correlation, I use OLS with an error structure that accounts for firm and year clustering. I also include year dummies to control year fixed effects. Petersen (2007) shows such estimation works well in panel datasets. To prevent a few outliers driving the result, I drop a month's observation if the average opening delay (time between 9:30 am and the first trade) is larger than 30 minutes.

Table 5 shows stock volatility is a decreasing function of the prior-openings. This result is consistent with the model's prediction that more opportunities to observe other peer stock prices lead to lower volatility. The size stratified result shows that the learning effect is not restricted to small, infrequently traded stocks. Rather, stocks with considerable size show a more significant learning effect. In order to show economic significance, table 5 also reports the regression result that uses minutes-to-first-trade as the main explanatory variable, instead of the prior-openings. The coefficient is negative 0.01. Because the difference between deciles is 0.1, this coefficient indicates the opening return volatility drops to the next lowest decile when a stock opens 10 minutes later.

Table 5

Analysis of opening return volatility

Model: $$Open\_vol_{im} = \pi \cdot p_{im} + \Gamma \cdot Z_{im} + \varepsilon_{im} \quad (7)$$

Dependent variable is monthly opening return variance. The opening return is calculated from two consecutive opening prices. p_{im} is prior-openings and Z_{im} is the matrix of control variables: market value, opening volume, Herfindahl index, and trading halt dummy. Prior-opening measures how many other stocks opened before the stock, and it is an inverse indicator of stock liquidity at open. To be included in the data, monthly average opening delay should be less or equal to 30 minutes. I run an OLS regression that is corrected for firm or year clustering and heteroskedastic error structure. Year dummies are used in the regression. In Panel A and B, coefficients are multiplied by 10^4 for visual convenience. T-values are in the parenthesis and significant variables in 1% level are marked with *.

Panel A: Regressions with different explanatory variables

	Model 1	Model 2	Model 3	Model 4
Prior-openings	-2.24*	-1.00*		-2.23*
	(-75.23)	(-39.83)		(-75.50)
Market Value	-770.63*		-242.73*	-769.89*
	(-69.58)		(26.46)	(-69.63)
Opening Volume	0.10*			
	(3.14)			
Herfindahl	864.63*			865.33*
	(12.98)			(12.99)
Trading Halt	0.26			
	(0.00)			
Observations	31391	31391	31391	31391
Adj. R-square	18.2%	5.4%	2.6%	18.2%

Panel B: Regressions by market value quartiles					
	All Stocks	Size (Min)	Size (2)	Size (3)	Size (Max)
Prior-openings	-2.24*	-1.63*	-2.21*	-2.43*	-2.45*
	(-75.23)	(-18.00)	(-25.38)	(-24.51)	(-18.99)
Market Value	-770.63*	-1214.14*	-830.63*	-558.76*	-658.11*
	(-69.58)	(-17.04)	(-6.04)	(-3.81)	(-12.34)
Opening Volume	0.10*	0.13	0.11	0.05	-0.14
	(3.14)	(2.32)	(1.72)	(1.27)	(-2.30)
Herfindahl	864.63*	1242.65*	354.60	403.03	1461.94*
	(12.98)	(5.51)	(1.54)	(1.38)	(4.91)
Trading Halt	0.26	-269.11	73.63	371.22	662.85
	(0.00)	(-2.05)	(0.58)	(2.44)	(2.79)
Observations	31391	7827	7860	7870	7846
Adj. R-square	18.2%	13.7%	17.9%	21.3%	26.7%

Panel C: Regressions with minutes-to-open variable					
	All Stocks	Size (Min)	Size (2)	Size (3)	Size (Max)
Minutes-to-Open	-0.010*	-0.009*	-0.012*	-0.015*	-0.013*
	(-45.53)	(-20.69)	(-22.46)	(-21.87)	(-9.91)
Market Value	-0.045*	-0.1263*	-0.073*	-0.056*	-0.017*
	(-45.75)	(-18.67)	(-5.21)	(-3.64)	(-3.02)
Opening Volume	0.000	0.000	0.000	0.000	-0.000*
	(1.49)	(2.09)	(1.27)	(0.86)	(-3.69)
Herfindahl	0.095*	0.107*	0.026	0.057	0.176*
	(13.47)	(4.78)	(1.05)	(1.89)	(5.40)
Trading Halt	0.03*	-0.007	0.055*	0.089*	0.103
	(3.88)	(0.56)	(4.02)	(4.98)	(3.86)
Observations	31391	7824	7856	7869	7846
Adj. R-square	9.8%	14.2%	12.5%	13.3%	16.4%

D. Learning Opportunity and Opening Friction

The analysis in table 5 does not differentiate two sources of stock volatility – the volatility from information itself and the volatility from pricing error. Amihud and Mendelson (1987) and Stoll and Whaley (1990) compares the opening return volatility with the losing return volatility to overcome this issue. Let opening return be the return between two consecutive opening prices, and the closing return be the return between two consecutive closing prices. Then take the monthly standard deviation of daily returns to get their volatility. The difference between the two volatilities (opening and closing), which is opening friction, represents the pricing error in opening prices, because two volatilities share the same 24-hour amount of information.

As a first step, I check whether the opening variance is also higher than the closing variance in 1997~2002 period for the NASDAQ data set. Table 6 confirms the opening variance is still 20% larger than the closing variance, as in Amihud and Mendelson (1987) and Stoll and Whaley (1990).

According to the model, the error in estimating the value of the information should be decreasing as opportunity to learn from other stock prices rises. The error in the opening price can be measured by the difference between the opening return volatility and the closing return volatility. So my third hypothesis is as follows:

Hypothesis 3: Opening friction decreases in the prior-openings.

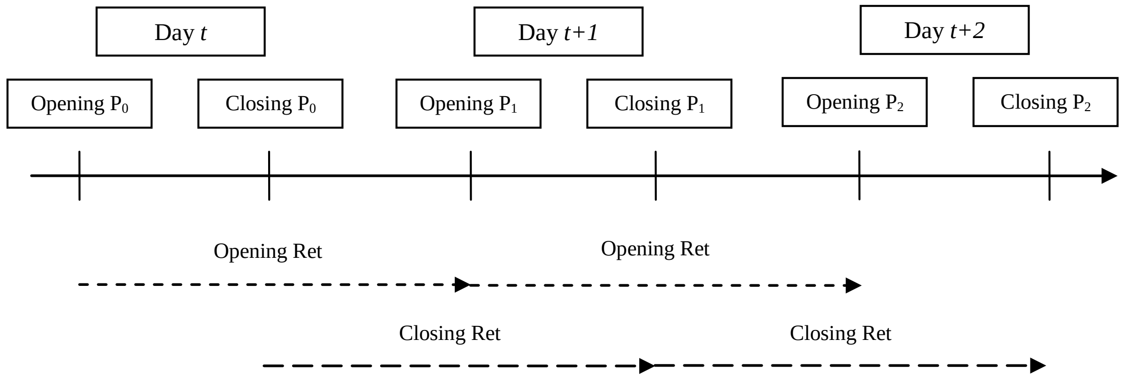


Figure 6. Comparing two returns.

Table 6

Size of opening friction

The stock returns are calculated from two consecutive opening prices (opening return) or two consecutive closing prices (closing return). Then I calculate monthly variation of two returns and divide opening return variation by closing return variation. To be included in the data, monthly average opening delay should be less or equal to 30 minutes. A number larger than 100% shows the opening variance is larger than the closing variance.

Market Value deciles	Opening friction = opening variance / closing variance	Observations
Min	125%	3137
2	128%	3141
3	123%	3139
4	134%	3140
5	121%	3138
6	115%	3142
7	116%	3139
8	114%	3138
9	116%	3141
Max	108%	3136
All stocks	120%	31398

Although opening friction in the literature is defined as the ratio between opening volatility and closing volatility, the ratio is not a good variable to use in OLS estimation. One problem with the ratio is that the measure is sensitive to the size of the denominator. This feature makes the variable highly skewed, and produces many outliers. So I subtract closing variance from opening variance and rank the difference into 11 categories. Then the deciles are modified as follows:

$$friction = (\mathrm{var}\_difference\_rank / 10) - 0.5$$

where *var_difference = opening return variance – closing return variance*

The structure of the test is the same as in the previous section, except that dependent variable is opening friction, instead of opening return volatility. We can see which variable is related to aggregate volatility and which variable is correlated with the pricing error only.

$$friction_{im} = \pi \cdot p_{im} + \Gamma \cdot Z_{im} + \varepsilon_{im} \qquad (8)$$

Table 7 shows the result of the estimation. I find opening friction is decreasing in the number of prior-openings. This result indicates that liquid stocks tend to have larger pricing errors. Comparing the result with the one in the previous section, opening volume and the Herfindahl index lose significance. Therefore, the size of the individual trade and industry structure are related to volatility of information arrivals rather, than volatility from pricing error. Panel C controls for the effect of bid-ask spread by using variances calculated from mid-quote returns. The mid-quote returns are derived from the quotes nearest to open and the quotes nearest to close. I find the result is robust, and adjusted R-square significantly increases.

I expect the size of opening friction will decrease in calendar time since there are increasing sources of to obtain price information available during the overnight periods. Specifically, there are more stocks listed in international markets. The volume of ECN is growing. There are more NASDAQ pre-opening quotes. Derivatives market is expanding to give more transaction information. Huang (2002) investigates the price information coming from ECNs. Cao, Ghysels, and Hatheway (2000) and Biais, Hillion, and Spatt (1999) show pre-opening quotes are closely related to subsequent transactions. Huang and Stoll (1994) argue investors can use futures price to discover the true price of a stock.

Table 7

Analysis of opening friction

Model: $friction_{im} = \pi \cdot p_{im} + \Gamma \cdot Z_{im} + \varepsilon_{im}$ (8)

To reduce the skewness and outlier problem, I rank the difference between opening return variance and closing return variance into deciles, and use it as dependent variable. p_{im} is prior-openings and Z_{im} is the matrix of control variables: market value, individual first volume, Herfindahl index, and trading halt dummy. Prior-opening measures how many other stocks opened before the stock, and it is an inverse indicator of stock liquidity at open. To be included in the data, monthly average opening delay should be less or equal to 30 minutes. I run an OLS regression that is corrected for firm or year clustering and heteroskedastic error structure. Year dummies are used in the regression. All coefficients are multiplied by 10^4 for visual convenience. T-values are in the parenthesis and significant variables in 1% level are marked with *.

Panel A: Regressions with different explanatory variables

	Model 1	Model 2	Model 3	Model 4
Prior-openings	-0.55*	-0.17*		-0.55*
	(-15.18)	(-5.69)		(-15.41)
Market Value	-231.25*		-102.59*	-232.13*
	(-16.71)		(-8.94)	(-16.82)
Opening Volume	-0.00			
	(-0.34)			
Herfindahl	14.20			
	(0.18)			
Trading Halt	84.43			
	(0.99)			
Observations	31391	31391	31391	31391
Adj. R-square	1.2%	0.2%	0.4%	1.2%

Panel B: Regressions by market value quartiles					
	All Stocks	Size (Min)	Size (2)	Size (3)	Size (Max)
Prior-openings	-0.55*	-0.71*	-0.47*	-0.65*	-0.27*
	(-15.18)	(-8.54)	(-6.81)	(-10.41)	(-3.70)
Market Value	-231.25*	-53.98	-441.09*	-381.75*	-84.83*
	(-16.71)	(-0.80)	(-4.09)	(-3.79)	(-2.66)
Opening Volume	-0.00	0.00	-0.16	0.03	-0.10
	(-0.34)	(0.62)	(-2.15)	(0.84)	(-1.57)
Herfindahl	14.20	81.85	40.03	-236.05	116.43
	(0.18)	(0.47)	(0.28)	(-1.66)	(0.70)
Trading Halt	84.43	17.40	240.91	26.56	87.90
	(0.99)	(0.11)	(1.65)	(0.16)	(0.36)
Observations	31391	7827	7860	7870	7846
Adj. R-square	1.2%	1.1%	1.3%	1.7%	1.3%

Panel C: Regressions using variances from mid-quote returns				
	Model 1	Model 2	Model 3	Model 4
Prior-openings	-1.34*	-1.68*		-1.34*
	(-24.50)	(-33.91)		(-24.70)
Market Value	250.48*		535.66*	249.53*
	(12.06)		(28.01)	(12.01)
Opening Volume	0.17*		0.03	0.17*
	(4.22)		(0.94)	(4.20)
Herfindahl	216.05			
	(1.85)			
Trading Halt	-72.67			
	(-0.58)			
Observations	24796	24796	24796	24796
Adj. R-square	10.2%	9.2%	6.5%	10.2%

F. Learning Effect vs. Stock Characteristics

One can argue that lower opening volatility and opening friction is a result of stock characteristics rather than a function of learning effect. To check this possibility, I study the case when a same stock's learning opportunity increases or decreases. For each stock, I calculate a stock's prior-openings change between two months. Then I compare it with the stock's volatility change during the same period. If my previous findings are mostly a result of stock characteristics, there will be little correlation between prior-openings change and volatility change. I use the same procedure to test correlation between prior-openings change and opening frictions change.

Table 8 shows the results. There is a significant negative correlation between prior-openings change and volatility change. The negative correlation is consistent with learning framework, because it means that a stock's volatility decreases when the stock gets more learning opportunity. The case of opening friction also shows a significant negative correlation. These results confirm that volatility and pricing error is a decreasing function of learning opportunity rather than a result of stock characteristics.

Table 8

Change of learning opportunity

For each stock, I take out the cases when a stock's prior-openings changes. Then I compare it with the stock's volatility change and opening friction change during the same period. I report correlation between the changes.

	Opening volatility change	Opening friction change
Correlation with prior-opening change	-0.226	-0.034
P-value of correlation coefficient	0.00	0.00

IV. Conclusion

This paper extends the literature on stocks volatility in two ways. First, I propose a price discovery model focusing on the volatility generated by the error in converting information into stock prices. While there are many models of volatility induced by private information, there are few models of volatility due to pricing error. This paper's model can be applied to pricing errors related phenomena, such as market opening frictions.

Second, I study the role of peer stocks on individual stock's volatility. The basic assumption is that information about one stock affects other stocks as well. Investors update their estimate of the value of a new piece of information by looking at the transaction prices of peer stocks. Thus, a stock where investors can observe more transactions of peer stocks should have less error in its price changes. The liquidity of peer stocks is beneficial, because investors would have a larger price data of peer stocks.

Third, less trading does not necessarily mean more pricing error. The learning effect can reduce pricing error of a stock without much trading. This implication questions the idea that a higher liquidity is always beneficial. Although scarce trading may delay price discovery process, excessive trading can also add noise to stock prices.

I verify the model's predictions using NASDAQ opening prices. In my dataset, opening prices contain a larger pricing error compared to closing prices. The higher pricing error can be explained by the model's prediction that a stock's pricing error increases when investors cannot observe the prices of peer stocks. I show that stock returns mean-revert to the cross sectional average after the market opening. This phenomenon is consistent with the model's implication that mean-reversion occurs when investors use peer stock price data to reduce pricing error in a particular stock. I also find that volatility and pricing error in opening prices decreases in a stock's learning opportunity. Stocks that open later than others tend to have less pricing error in their opening prices and have a weaker tendency to mean-revert.

References

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

Amihud, Yakov and Mendelson, Heim, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223-249.

Amihud, Yakov and Mendelson, Heim, 1987, Trading mechanisms and stock returns: An empirical investigation, *Journal of Finance* 42, 533-553.

Amihud, Yakov and Mendelson, Heim, 1991, Volatility, efficiency, and trading: Evidence from the Japanese stock market, *Journal of Finance* 46, 1765-1789.

Acharya, Viral V. and Pedersen, Lasse Heje, 2005, Asset pricing with liquidity risk, *Journal of Financial Economics* 77, 375-410.

Anushman, V. Ravi, Chordia, Tarun and Subrahmanyam, Avanidhar (2001), Trading activity and expected stock returns, *Journal of Financial Economics* 59, 3-32.

Bernard, Victor L. and Thomas, Jacob K., 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics* 13, 305-340.

Biais, Bruno, Hillion, Pierre, and Spatt, Chester, 1999, Price discovery and learning during the pre-opening period in the Paris Bourse, *Journal of Political Economy* 107, 1218-1248.

Boehme, Rodney D., Danielsen, Bartley R., Sorescu, Sorin M., 2006, Short sale constraints, differences of opinion, and overvaluation, *Journal of Financial and Quantitative Analysis 41*, 455-487.

Brennan, Michael J., Chordia, Tarun and Subrahmanyam, Avanidhar, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics* 49, 345-373.

Brennan, Michael J., Jegadeesh, Narasimhan, and Swaminathan, Bhaskaran, 1993, Investment analysis and the adjustment of stock prices to common information, *Review of Financial Studies* 6, 799-824.

Brennan, Michael J., Subrahmanyam, Avanidhar, 1995, Investment analysis and price formation in securities markets, *Journal of Financial Economics* 38, 361-382.

Brennan, Michael J., Subrahmanyam, Avanidhar, 1996, Market microstructure and asset pricing: On the compensation for illiquidity in stock returns, *Journal of Financial Economics* 41, 441-464.

Brooks, Raymond M. and Su, Tie, 1997, A simple cost reduction strategy for small liquidity traders: Trade at the opening, *Journal of Financial and Quantitative Analysis* 32, 525-541.

Campbell, John Y., Grossman, Sanford J., Wang, Jiang, 1993, Trading volume and serial correlation in stock returns, *Quarterly Journal of Economics* 108, 905-939.

Cao, Charles, Ghysels, Eric and Hatheway, Frank, 2000, Price discovery without trading: Evidence from the NASDAQ preopening, *Journal of Finance* 55, 1339-1365.

Chan, Kalok, 1993, Imperfect information and cross-autocorrelation among stock prices, *Journal of Finance* 48, 1211-1230.

Chan, K. C., Christie, William G., and Schultz, Paul H., 1995, Market structure and intraday pattern of bid-ask spreads for NASDAQ securities, *Journal of Business* 68, 35-60.

Chordia, Tarun and Subrahmanyam, Avanidhar, 2004, Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Economics* 72, 485-518.

Chordia, Tarun and Swaminathan, Bhaskaran, 2000, Trading volume and cross-autocorrelations in stock returns, *Journal of Finance* 55, 913-935.

Clinch, Greg J. and Sinclair, Norman A, 1987, Intra-industry information releases: A recursive systems approach, *Journal of Accounting and Economics* 9, 89-106.

Datar, Vinay T., Naik, Narayan Y. and Radcliffe, Robert, 1998, Liquidity and stock returns: an alternative test, *Journal of Financial Markets* 1, 203-219.

Easley, David and O'Hara, Maureen, 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 577-605.

Easley, David and O'Hara, Maureen, 2004, Information and the cost of capital, *Journal of Finance* 59, 1553-1583.

Easley, David and O'Hara, Maureen, 2003, Microstructure and asset pricing (Ch. 17), *Handbook of the Economics of Finance*, Editors: George M. Constantinides, Milton Harris and Rene M. Stulz, Elsevier Press

Easley, David, Hvidkjaer, Soeren, and O'Hara, Maureen, 2002, Is information risk a determinant of asset returns?, *Journal of Finance* 57, 2185-2221.

Evans, Martin D.D., Lyons, Richard K.., 2002, Order flow and exchange rate dynamics, *Journal of Political Economy* 110, 170‒180.

Evans, Martin D.D., Lyons, Richard K.., 2008, How is macro news transmitted to exchange rates?, *Journal of Financial Economics* 88, 26‒50.

Forster, Margaret M. and George, Thomas J., 1996, Pricing errors at the NYSE open and close: Evidence from internationally cross-listed stocks, *Journal of Financial Intermediation* 5, 95-126.

French, Keneth R. and Roll, Richard, 1986, Stock return variances: the arrival of information and the reaction of traders, *Journal of Financial Economics* 17, 5-26.

Gervais, Simon, Kaniel Ron and Mingelgrin, Dan H., 2001, The high volume return premium, *Journal of Finance* 56, 877-919.

Han, Jerry C. Y. and Wild, John J., 1990, Unexpected earnings and intra-industry information transfers: Further evidence, *Journal of Accounting Research* 28, 211-219.

Harris, Milton and Raviv, Artur, 1993, Differences of opinion make a horse race, *Review of Financial Studies 6*, 473-506.

Hong, Harrison and Stein, Jeremy C., 2003, Differences of opinion, short-sales constraints, and market crashes, *Review of Financial Studies 16*, 487-525.

Huang, Roger D., 2002, The quality of ECN and NASDAQ market maker quotes, *Journal of Finance* 57, 1285-1319.

Huang, Roger D. and Stoll, Hans R., 1994, Market microstructure and stock return predictions, *Review of Financial Studies* 7, 179-213.

Karolyi, Andrew G., and Stulz, Rene M., 2003, Are financial assets priced locally or globally? (Ch. 16), *Handbook of the Economics of Finance*, Editors: George M. Constantinides, Milton Harris and Rene M. Stulz, Elsevier Press

Kim, Oliver and Verrecchia, Robert E., 1994, Market liquidity and volume around earnings announcements, *Journal of Accounting and Economics* 17, 41-67.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-1335.

Lo, Andrew, and MacKinlay, Craig, 1990, When are contrarian profits due to stock market overreaction?, *Review of Financial Studies* 3, 175-205.

Madhavan, Ananth and Panchapagesan, Venkatesh, 2000, Price discovery in auction markets: A look inside the black box, *Review of Financial Studies* 13, 627-658.

Madhavan, Ananth and Sofianos, Geroge, 1998, An empirical analysis of NYSE specialist trading, *Journal of Financial Economics* 48, 189-210.

Madhavan, Ananth, Richardson Matthew, and Roomans, Mark, 1997, Why do security prices change? A transaction-level analysis of NYSE stocks, *Review of Financial Studies* 10, 1035-1064.

Masulis, Ronald W., and Shivakumar, Lakshmanan, 2002, Does market structure affect the immediacy of stock price response to news?, *Journal of Financial and Quantitative Analysis* 37, 617-648.

Mendenhall, Richard R., 2004, Arbitrage risk and post-earnings-announcement drift, *Journal of Business* 77, 875 – 894.

O'Hara, Maureen, 2003, Liquidity and price discovery, *Journal of Finance* 58, 1335-1354.

Pastor, Lubos and Stambaugh, Robert F., 2001, Liquidity premium and expected stock returns, *Journal of Political Economy 111*, 642-685.

Pasquariello, Paolo and Vega, Clara, 2008, Strategic cross trading in US market, University of Michigan working paper.

Petersen, Mitchell A., 2007, Estimating standard errors in finance panel data sets: comparing approaches, *forthcoming Review of Financial Studies.*

Stoll, Hans R., 2000, Friction, *Journal of Finance* 55, 1479-1514.

Stoll, Hans R., and Whaley, Robert E., 1990, Stock market structure and volatility, *Review of Financial Studies* 3, 37-71.

Veldkamp, Laura L., 2006, Information markets and the comovement of asset prices, *Review of Economic Studies* 73, 823-845.

Appendix: Proof of Results

Result 1: A stock's price change after an information event's arrival converges to the average price change of its peer stocks.

Proof 1: Assume trading cost is small compared to prices. The average of stock price changes is:

$$\frac{1}{n}\sum_{i=1}^{n}(v+e_i) \quad \text{, where n is the number of the stocks in the peers}$$

Now think of a strategy that sells stocks that moved higher than sample average and buys stocks that moved lower than sample average. Stock price change eventually converges to v. The profit of this strategy for a stock j that moved higher than sample average is:

$$v + e_j - v \qquad \text{if} \qquad v + e_j > \frac{1}{n}\sum_{i=1}^{n}(v+e_i)$$

This equation can be simplified as:

$$e_j \qquad \text{if} \qquad e_j > \frac{1}{n}\sum_{i=1}^{n}(e_i)$$

The sample average of errors, $\frac{1}{n}\sum_{i=1}^{n}(e_i)$, is a normally distributed variable because e_i is normally distributed. Taking this strategy continuously over time, sample average of the errors converges to 0 in probability by Law of Large Numbers. Then e_j becomes the random draw from truncated distribution that has minimum 0 on average. Because the expected value of the minimum is 0, expected value of e_j is positive. By the same logic, profit from a stock moved lower than sample average is also positive.

Rational investors would take this opportunity, and by combining several stocks, they can create an arbitrage position. In no arbitrage market, these trading activities should bring stock prices to industry sample average. ■

Result 2: A stock's price change after the information arrival becomes more accurate, the larger is the sample of previously traded peer stocks.

Proof 2: I show the accuracy of sample average depends on the number of traded stocks. Let there be *l* traded stocks. The variance of this industry average is:

$$VAR(\frac{1}{l}\sum_{i=1}^{l}(v+e_i)) = E(\frac{1}{l}\sum_{i=1}^{l}(v+e_i)-v)^2 = E(\bar{v}-v)^2$$

Expansion of above equation gives:

$$VAR(\bar{v}) = E(\bar{v}-v)^2 = E(\frac{1}{l}(\sum_{i=1}^{l}v+e_i)-v)^2 = E(\frac{1}{l^2}\sum_{i=1}^{l}(e_i)^2)+\sum\sum_{i\neq j}\frac{1}{l^2}E(e_ie_j) = E(\frac{1}{l^2}\sum_{i=1}^{l}(e_i)^2)$$

Since error term *e* is independent each other, the covariance term is 0.

$$VAR(\bar{v}) = E(\frac{1}{l^2}\sum_{i=1}^{l}(e_i)^2) = E(\frac{1}{l}\cdot\frac{1}{l}\sum_{i=1}^{l}(e_i)^2) = \frac{1}{l}E(\frac{1}{l}\sum_{i=1}^{l}(e_i)^2) = \frac{VAR(e)}{l}$$

As *l* gets larger, variance of sample average gets lower. A stock trading after observing this average can estimate the true value with lower error. ■

CHAPTER 2

IS ORDER IMBALANCE RELATED TO INFORMATION?

I. Introduction

Order imbalance is the signed volume that measures direction and degree of buying or selling pressure. Such trading pressure can signal investors' interest in a particular stock, and order imbalance may predict future stock returns (Chordia and Subrahmanyam 2004). Market participants give considerable effort to estimate order imbalance, and investigate what information is underlying it. It is easy to assume that order imbalance is an indicator of private information, because random orders would be unlikely to have a consistent direction. Practitioners are already using the idea that order imbalance shows underlying information. Wall Street Journal posts daily order imbalance under the name 'Buying on Weakness' or 'Selling on Strength'.

However, some papers that use order imbalance data find weak correlation between order imbalance and information arrivals. Andrade, Chang, and Seasholes (2008) argue significant amount of trading imbalances are uninformed trading, but the imbalances affect stock returns. Easley, Engle, O'Hara, and Wu (2008) show order imbalance is not particularly good at capturing the presence of informed trading. Lee (1992) and Trueman, Wong, and Zhang (2003) find investors do not place orders according to announced information. Even if a large, positive earnings surprise occurs, for example, order imbalance is not always positive. On the other hand, order imbalance exhibits consistent patterns, which suggest the presence of rational trading activity. Chordia and Subrahmanyam (2004) find daily order imbalance is serially correlated, and past order imbalance predicts future returns. They explain this finding within the framework of Kyle (1985). Kyle (1985) models the trading strategy of a rational investor, who has private information. The investor splits her order into a number of small orders to hide the private information from public, and such order fragmentation makes order imbalance serially correlated. Thus, under his framework, order imbalance reflects private information. Easley, O'Hara, and Paperman (1996), and Easley, Kiefer, and O'Hara (1997) also argue private positive or negative information generates buying or selling pressure respectively.

The question of whether order imbalance is related to information is important because order imbalance has close link with asset prices. Chordia and Subrahmanyam (2004) and Chordia, Roll, and Subrahmanyam (2008) find a stock's order imbalance is positively correlated with its future return. Bollen and Whaley (2004) shows option values are affected by buying or selling pressure. Brandt and Kavajecz (2004) find order imbalance is a major source of bond yield fluctuations. The connection between order imbalance and asset prices can contain different meanings depending on the origin of the order imbalance. The relation between order imbalance and return, and the relation between return and information has been actively investigated over the last 10 years, but the relation between order imbalance and information has not received much attention yet.

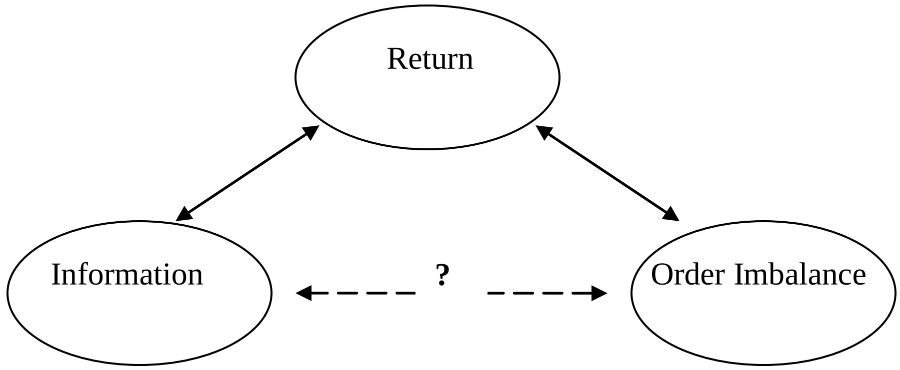


Figure 1. Triangular relation among return, information, and order imbalance

To shed light on the triangular relation among returns, order imbalance, and information, this paper studies the link between order imbalance and information. I focus on stock earnings surprises. An earnings surprise is defined as the difference between actual earnings and average analyst expectations, which measures the size of the informational shock by the market. Numerous papers including Bernard and Thomas (1990), Bhushan (1994), Bartov, Radhakrishnan, and Krinsky (2000), and Mendenhall (2004) use earnings surprises to proxy the size of new information shocks. My test is has two parts: First, I test whether order imbalance acts as an indicator of private information. If order imbalance is generated by private information, it may well predict the forthcoming major information event like an earnings surprise. The second part of the test is whether order imbalance is correlated with public information. If order imbalance reflects a public announcement, a positive earnings surprise would be accompanied by positive order imbalance and vice versa. I construct empirical hypotheses from the existing theories on order imbalance. Because these theories are based on the assumptions of market efficiency and rational investors, my tests can also be used as evidence for/against market efficiency and rational investors.

I find following relation between order imbalance and information:

(1) Order imbalance has no reliable predictive power for a forthcoming information event. Order imbalance before an earnings announcement date has in general insignificant correlations with earnings surprise. Moreover, order imbalance does a poor job in predicting whether there will be a positive earnings surprise or not. This result is disappointing to investors who want to use order imbalances to aid their investment

decisions, but it is consistent with weak market efficiency hypothesis which tells that past data cannot predict subsequent stock returns.

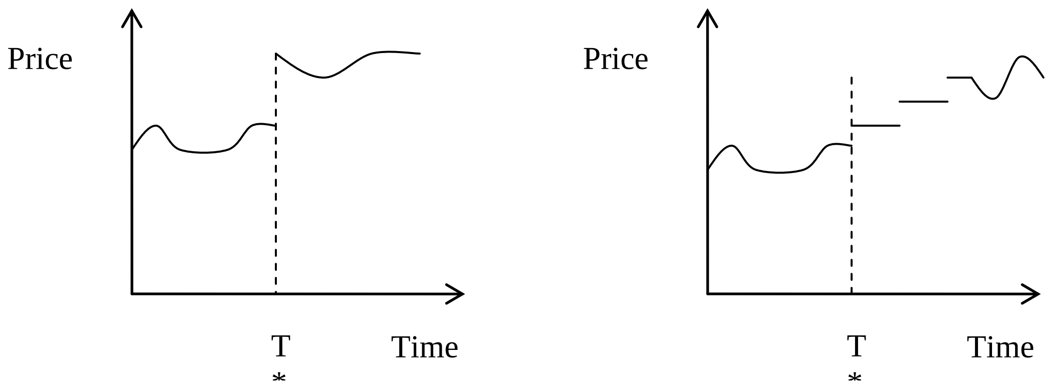
(2) Order imbalances at earnings announcements act independently of earnings surprise. This result can be explained in market efficiency framework. When information arrives, market makers quickly adjust their quotes according to the information. If they move faster than other traders, the adjusted quote level reflects the announcement and there is no need to place additional buy or sell orders as a result of the public announcement. Thus, order imbalance can be independent of the announcement. The weak correlation between order imbalance and public information arrival can be evidence of market makers' fast quote movements. Fleming and Remolona (1999) find in the Treasury market that quote adjustments precede trades when major economic data is publicly released.

(3) Past earnings announcements have a positive correlation with order imbalance. After earnings data are released to the public, order imbalance starts to mimic the earnings surprise for several days. This pattern continues for several days after the public announcement. Unlike (1) or (2), this result cannot be explained by the market efficiency hypothesis. Why should investors place orders according to the past information that has no value in an efficient market? The correlation between order imbalances and past earnings announcements can not be ignored, because I find stock prices also react to order imbalances.

My three results show that the relationship between order imbalance and information is complicated. Order imbalance has weak correlation with information before announcement dates, which is consistent with the efficient market hypothesis. However, order imbalance has correlation with past information, which can be explained by investors' irrational behaviors. Without proper framework that can explain the behaviors of order imbalance, using order imbalance variable in investment decisions can be a risky idea.

II. Hypotheses on Order Imbalance and Information

I start by constructing hypotheses on relation between order imbalance and information, based on the assumption of efficient markets. Following standard procedure of Lee and Ready (1992), order imbalance is calculated by subtracting the trades in bid side (selling pressure) from the trades in ask side (buying pressure). There will be relatively more bid side transactions when current quotes are above market's consensus price, and vice versa. Therefore, order imbalance captures the dispersion of opinion between market makers (who sets quotes) and other investors. Even if there is highly positive information, order imbalance can be negative if market makers post their quotes above the consensus price. The following figure compares two cases: when market makers change their quotes according to positive information and when they do not. Order imbalance will not reflect the value of public information if market makers change their quotes quickly.



Stock A – fast moving quotes (No order imbalance)

Stock B – slow or not moving quotes (High order imbalance)

Figure 2. Quote speed and order imbalance.

A. Order imbalance and private information

Due to its dependence on current quote level, order imbalances follow the direction of information when market makers do not instantly adjust their quotes. When market makers do not know the information that some other investors do, the informed investors would trade based on their *private* information, and their trade will generate order imbalance. If the private information is to be announced later, the order imbalance pattern would predict the forthcoming announcement.

Kyle (1985) is one of the most cited models of relation between private information and order imbalance.[14] He uses game theory to explain the relation between order imbalance and information when the information is not public. In other words, information is private in his model. His model shows order imbalance is a function of information, market depth, and price level.

$$\Delta\tilde{x}_n = \beta_n(\tilde{v} - \tilde{p}_{n-1})\Delta t_n \qquad (1)$$

,where β is a decreasing function of market depth λ.

Δx is order placement of informed investors, β is a decreasing function of Kyle's lambda λ (market depth), v is value of private information, p is stock's price level, t is the time left until information is publicly announced, and there are N trades before the information is announced.

β is decreasing in λ, so I substitute β with $f^{-1}(\lambda)$. $f(\lambda)$ is an increasing function of λ. Rewriting equation (1), I get:

[14] Back and Bruch (2004) extend Kyle (1985) model to show that the model's implications also hold in continuous trading. Chordia and Subrahmanyam (2004) uses Kyle (1985) model to explain serial correlation of order imbalance.

$$\tilde{v} = \frac{\Delta\tilde{x}_n}{\Delta t_n} \cdot f(\lambda_n) + \tilde{p}_{n-1} \qquad (2)$$

Equation (2) states that value of private information is related to informed order placement *Δx*, market depth *λ*, and previous price *p*. Equation (2) is also empirically testable, because all the variables can be obtained from stock market. Although informed order placement is not observable to the public, order imbalance will be proportional to *Δx*, because other orders have no direction. All investors observe aggregate order flow *Δx* + *η*, where *η* is order flow with zero mean. Hence, order imbalance is an unbiased estimator of informed order placement *Δx*. Using standard regression techniques, one can filter out the effect of *η*, because the mean of *η* is zero. Equation (3) substitutes informed order placement *Δx* with observed order imbalance *Δx* + *η*.

$$\begin{aligned} \tilde{v} &= \frac{(\Delta\tilde{x}_n + \eta)}{\Delta t_n} \cdot f(\lambda_n) + \tilde{p}_{n-1} \\ &= \frac{(\Delta\tilde{x}_n)}{\Delta t_n} \cdot f(\lambda_n) + \tilde{p}_{n-1} + \eta \cdot \frac{f(\lambda_n)}{\Delta t_n} \qquad (3) \end{aligned}$$

By taking average of equation (3), we get:

$$\begin{aligned} \overline{\tilde{v}} &= \frac{(\Delta\overline{\tilde{x}}_n)}{\Delta t_n} \cdot f(\lambda_n) + \overline{\tilde{p}}_{n-1} + \overline{\eta} \cdot \frac{f(\lambda_n)}{\Delta t_n} \\ &= \frac{(\Delta\overline{\tilde{x}}_n)}{\Delta t_n} \cdot f(\lambda_n) + \overline{\tilde{p}}_{n-1} \qquad (4) \end{aligned}$$

Equation (4) shows that the value of information is increasing in previous price level and the product of order imbalance and Kyle's lambda. Because Kyle's lambda and previous price level can be estimated using past data, investors can detect the value of information using contemporaneous order imbalance. The time frame Δt can be substituted with 1 when one uses same time frame to estimate order imbalance (ex. daily order imbalance).

The validity of equation (4) can be tested using regression equation (5):

For stock *i*, information announcement at day *t*,

$$v_{it} = \alpha + \delta_1 \cdot (OI_{it-j}) \cdot (\lambda_{it-j}) + \delta_2 \cdot r_{i,t-j-1} + \varepsilon_{it} \qquad (5)$$

,where *j* indicates the time between the forthcoming information announcement and current order imbalance.

Equation (5) substitutes informed order flow with order imbalance and uses stock return in place of price level. δ_1 and δ_2 are variables to estimate. I use stock return in equation (5) because price level in Kyle's model does not account for price level variation across stocks, but represents the amount of information captured by price. In a cross sectional regression, return is a more appropriate choice to measure the effect.

Following Kyle (1985), my first empirical hypothesis is that δ_1 in equation (5) is positive and significant.

H1: Order imbalance is positively correlated with the value of forthcoming information.

B. Order imbalance and public information

Although order imbalance can be a function of *private* information, order imbalance reacts differently to *public* information. When there is a public announcement, market makers also know the information. Order imbalance may still reflect the direction of a public information announcement, when market makers fail to adjust their quotes quickly enough to the announcement. However, in an efficient market, quote adjustment should be faster than any trades.

Suppose there is a positive announcement from a company. If market makers do not change their quote quickly, they would sell their stocks at discount and some investors may make profit from the public announcement. This violates semi-strong efficiency, which requires a public announcement to have no investment value. Hence, in a semi-strong efficient market, quotes should move before any trade comes in.[15] This independence argument may seem to contradict equation (2), which states that order imbalance has some connection to the value of upcoming information. However, both arguments are based on the assumption that market is efficient. It is the timing of announcement or the nature of information that changes the behavior of order imbalance. In efficient market, order imbalance may predict forthcoming information, but reflect current announcement.

Equation (2) can be modified to express public information case. When price instantly adjusts to the value of the announced information, $v = p$ in equation (2).

$$\tilde{v} = \frac{\Delta\tilde{x}_n}{\Delta t_n} \cdot f(\lambda_n) + \tilde{p}_{n-1}$$

Then informed order placement Δx becomes zero.

$$\frac{\Delta\tilde{x}_n}{\Delta t_n} \cdot f(\lambda_n) = 0$$

[15] Fama (1970) explains the definition of semi-strong efficiency. In such market, no investor should be able to profit from public information.

This argument implies that in any semi-strong efficient market, public information will be converted to price in a 2-step procedure. In step one, public information arrives and quotes first adjust according to the information. Trades occur in the second step to trade based on the quote level. Such 2-step procedure means that without any help from trading, investors can successfully estimate the unbiased price from public information. Fleming and Remolona (1999) find such 2-step pattern in Treasury Bill market.

So my second hypothesis is:

H2: Order imbalance is independent of contemporaneous public announcement.

If hypothesis 2 is valid, order imbalance should be also independent of past announcement. Unless information is serially correlated, order imbalance would not reflect past information.[16]

H3: Order imbalance is independent of past announcement.

I will test these 3 hypotheses in the next section.

III. Data and Method

I use Trade and Quote (TAQ) data for ordinary common shares from 1996 to 2004 to construct order imbalance data. The construction method is in appendix, and it closely follows the method of Chordia, Roll, and Subrahmanyam (2002). Their method is based on Lee and Ready (1992), but it imposes additional filters to reduce problems from scarce trading. I report the results using order imbalance of shares, dollars, and number of trades.

[16] Chordia and Shivakumar (2006) find some information can be serially correlated, such as earnings information.

In my dataset, order imbalance of shares and order imbalance of dollars have 99% positive correlation, while order imbalance of trades has a lower 83% positive correlation with other two measures. I normalize order imbalance measures by dividing it by daily aggregate share volume, dollar volume, or number of trades respectively. This setting allows cross sectional comparisons across stocks.

As in Chordia and Subrahmanyam (2004), I use mid-quote stock returns to take out the effect of bid-ask bounce. The mid-quotes data comes from Market Microstructure Database in Vanderbilt University. The database contains average trade weighted bid price and ask price during a day. I take mid-quote prices from the bid price and the ask price, and calculate continuously compounded return using two consecutive mid-quote prices (taking natural logarithm of the ratio).

In order to measure the value of information, I choose earnings announcement event. For each earnings announcement, there are analyst forecasts for earnings per share. The earnings surprise, which is the difference between the forecasted earnings (analyst consensus) and the actual earnings, represents the value of information to the stock market. Earnings surprise data comes from IBES database. I use quarterly earnings announcements. I use the most recent earnings per share (EPS) forecasts for each analyst, and if the forecast is more than 90 days old, I drop the forecast. I require stocks to have more than 5 recent forecasts. Mendenhall (2004) and other earnings surprise related papers define earnings surprise as follows: Earnings surprise is difference between actual earnings and average analyst EPS forecasts, divided by standard deviation of the forecasts.

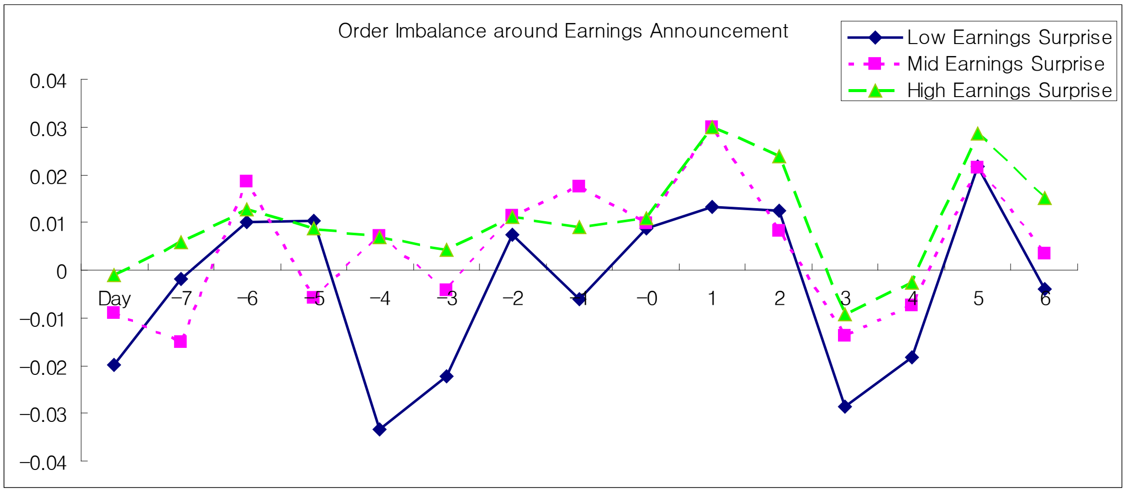
$$Surprise_{i,q} = \frac{Actual\_Earnings_{i,q} - \overline{Exp\_Earnings_{i,q}}}{STD(Exp\_Earnings_{i,q})}$$

where i is one firm and q is one quarter.

Bernard and Thomas (1990) use ranks of the earnings surprise to account for non linearity and outlier problem. As Mendenhall (2004) suggests, I rank earnings surprise into 11 ranks and then divide 11 and subtract 0.5 from the variable. The ranked earnings surprise variable has its mean around 0, and 0.1 is the difference between two close ranks. I call this variable surprise rank.

Stoll (2000) shows Kyle's lambda λ can be measured as the sensitivity of stock return to order imbalance. For every trading day, I use 250 prior business days of order imbalance data to estimate Kyle's lambda. I regress stock return on order imbalance, and Kyle's lambda is the coefficient. Because of the lambda estimation, the final data set has time period from 1997 to 2004. After all the adjustments, I have 2,854 earnings announcements in my dataset.

The following figures show average order imbalance and mid-quote return by 3 ranks of earnings surprise.



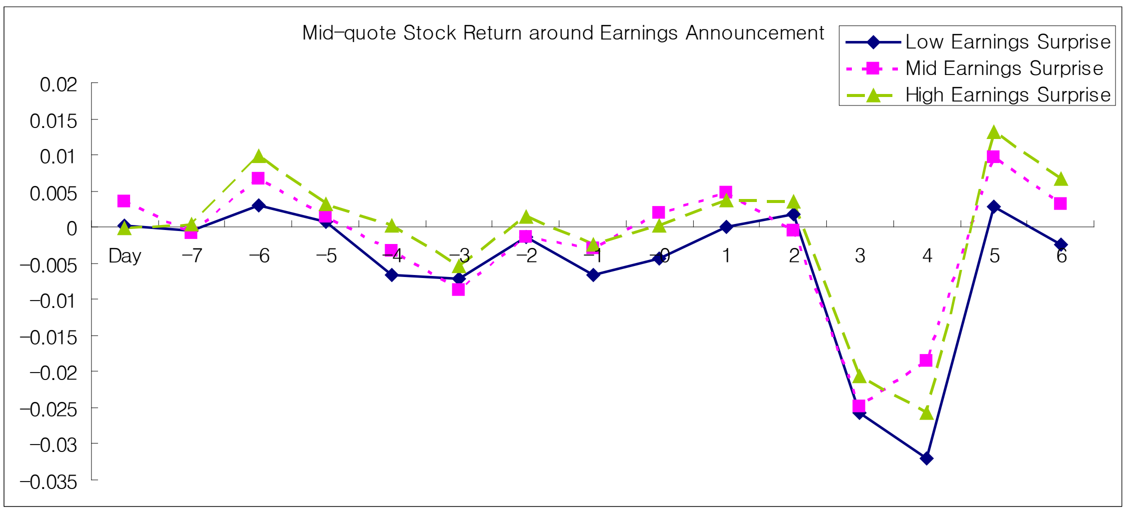


Figure 3. Order imbalance and mid-quote return around earnings announcements

IV. Empirical Tests

A. Order imbalance and private information

First I test the predictive power of order imbalance using equation (5).

$$v_{it} = \alpha + \delta_1 \cdot (OI_{it-j}) \cdot (\lambda_{it-j}) + \delta_2 \cdot r_{i,t-j-1} + \varepsilon_{it} \qquad (5)$$

If coefficient δ_1 is statistically significant and economically meaningful, order imbalance would be a useful variable to predict upcoming earnings announcement. Regression method is OLS with clustering and heteroskedasticity controlled error structure. I control for clustering by firm, year, and quarter. Petersen (2007) shows such correction yields a consistent estimation for panel data sets.

Table 1 shows δ_1 is mostly insignificant. This result indicates order imbalance has almost no forecasting power for future announcement. δ_1 is only significant at 2 days before an announcement. At that point, order imbalance has little value for investors because the order imbalance is measured after market close. Investors have only 1 business day to trade based on the order imbalance, and by that time, there can be a quite a lot of information leakage such as CEOs announcing hints about their earnings. Even if the leakage does not occur, order imbalance is not useful for investment decisions because of its lack of significance. Adjusted R-square is lower than 1%, and the size of the coefficient is less than 1. Given that order imbalance has range between -1 ~ +1 and my Kyle's lambda variable has average value of 0.12, the small size of coefficient implies that order imbalance can explain at most one-tenth of difference in earnings surprise rank.

Table 1 – Order imbalance and private information

$$v_{it} = \alpha + \delta_1 \cdot (OI_{it-j}) \cdot (\lambda_{it-j}) + \delta_2 \cdot r_{i,t-j-1} + \varepsilon_{it} \qquad (5)$$

Earnings surprise measure *v* is derived from the difference between actual earnings and average analyst forecasts. I rank the earnings surprise into 11 ranks, divide it by 10, and subtract 0.5. *OI* is order imbalance and r_{t-1} is previous day's mid-quote stock return. For every trading day, I use 250 prior business days of order imbalance data to estimate Kyle's lambda. I regress stock return on order imbalance, and Kyle's lambda is the coefficient. Regression method for equation (3) is OLS with clustering and heteroskedasticity controlled error structure. P-values are in the parenthesis. Coefficients significant in 1%, 5%, and 10% level are marked with a, b, and c.

Panel A: Regression with order imbalance of shares

Days before announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	0.939[b] (0.01)	0.235 (0.40)	1578	0.5%
3	0.322 (0.39)	0.291[c] (0.07)	1629	0.2%
4	-0.646[c] (0.06)	0.188 (0.20)	1924	0.2%
5	0.681 (0.18)	0.087 (0.75)	1601	0.1%
6	0.235 (0.57)	0.065 (0.68)	2209	0.0%
7	0.625[c] (0.07)	0.067 (0.61)	2891	0.1%

Panel B: Regression with order imbalance of dollars

Days before announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	0.940[b] (0.01)	0.235 (0.40)	1578	0.5%
3	0.316 (0.40)	0.291[c] (0.07)	1629	0.2%
4	-0.649[c] (0.06)	0.189 (0.20)	1924	0.2%
5	0.681 (0.19)	0.088 (0.75)	1601	0.1%
6	0.235 (0.57)	0.066 (0.68)	2209	0.0%
7	0.614[c] (0.08)	0.068 (0.61)	2891	0.1%

Panel C: Regression with order imbalance of trades

Days before announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	0.841[a] (0.00)	0.273 (0.33)	1578	0.6%
3	0.688[b] (0.01)	0.282[c] (0.08)	1629	0.6%
4	0.141 (0.61)	0.181 (0.22)	1924	0.1%
5	0.552 (0.11)	0.102 (0.71)	1601	0.2%
6	0.303 (0.29)	0.073 (0.64)	2209	0.1%
7	0.320 (0.14)	0.087 (0.51)	2891	0.1%

In table 2, I run regressions after including contemporaneous order imbalance in equation (5). As in Chordia and Subrahmanyam (2004), order imbalances are serially correlated, and so having contemporaneous order imbalance in the test equation may reveal additional explanatory power of order imbalance. Contemporaneous order imbalance is average order imbalance in (-1, +1) day window around an announcement date.[17]

$$v_{it} = \alpha + \delta_1 \cdot (OI_{it-j}) \cdot (\lambda_{it-j}) + \delta_2 \cdot r_{i,t-j-1} + \delta_3 \cdot (\overline{OI_{it}}) + \varepsilon_{it} \qquad (5)'$$

Table 2 shows that order imbalance does not gain much explanatory power by adding contemporaneous order imbalance. Again I find that δ_1 is insignificant in most cases, R-squares are low, and coefficients are too small to be economically meaningful.

The low significance of δ_1 may come from strategic trading of informed investors. For example, Back and Baruch (2004) argue informed traders can place fake orders to disguise their trading. My result points to three possibilities. 1) Informed investors are very good at hiding their trades, 2) order imbalance is generated by some other factor than information, or 3) the model specification of equation (5) is incorrect.

[17] Even if I use only order imbalances at announcement dates, I get qualitatively similar results.

Table 2 – Regressions with contemporaneous order imbalance

$$v_{it} = \alpha + \delta_1 \cdot (OI_{it-j}) \cdot (\lambda_{it-j}) + \delta_2 \cdot r_{i,t-j-1} + \delta_3 \cdot (\overline{OI_{it}}) + \varepsilon_{it} \qquad (5)'$$

Contemporaneous order imbalance $\overline{OI_{it}}$ is the average order imbalance in (-1, +1) day window around an earnings announcement. Regression method for equation (3) is OLS with clustering and heteroskedasticity controlled error structure. P-values are in the parenthesis. Coefficients significant in 1%, 5%, and 10% level are marked with a, b, and c.

Panel A: Regression with order imbalance of shares, with contemporaneous order imbalances					
Days before announcement	δ_1	δ_2	δ_3	*Observations*	*Adj. R-square*
2	0.881[b] (0.02)	0.224 (0.42)	0.064 (0.32)	1578	0.5%
3	0.222 (0.56)	0.291[c] (0.07)	0.096 (0.12)	1629	0.4%
4	-0.684[b] (0.05)	0.187 (0.21)	0.062 (0.29)	1924	0.3%
5	0.679 (0.18)	0.088 (0.75)	0.003 (0.96)	1601	0.1%
6	0.217 (0.60)	0.068 (0.67)	0.025 (0.65)	2209	0.0%
7	0.593[c] (0.09)	0.075 (0.57)	0.047 (0.33)	2891	0.2%

Panel B: Regression with order imbalance of dollars, with contemporaneous order imbalances					
Days before announcement	δ_1	δ_2	δ_3	*Observations*	*Adj. R-square*
2	0.789[b] (0.02)	0.206 (0.42)	0.048 (0.42)	1578	0.5%
3	0.291 (0.40)	0.232[c] (0.10)	0.077 (0.17)	1629	0.3%
4	-0.606[b] (0.05)	0.186 (0.17)	0.049 (0.35)	1924	0.3%
5	0.632 (0.17)	0.096 (0.70)	-0.000 (0.99)	1601	0.1%
6	0.148 (0.69)	0.063 (0.65)	0.012 (0.82)	2209	0.0%
7	0.551[c] (0.08)	0.030 (0.80)	0.033 (0.45)	2891	0.1%

Panel C: Regression with order imbalance of trades, with contemporaneous order imbalances					
Days before announcement	δ_1	δ_2	δ_3	*Observations*	*Adj. R-square*
2	-0.043 (0.91)	0.26 (0.37)	0.144 (0.00)	1578	0.6%
3	0.056 (0.89)	0.257[c] (0.07)	0.138[a] (0.01)	1629	0.7%
4	-0.355 (0.30)	0.173 (0.20)	0.068 (0.17)	1924	0.2%
5	-0.055 (0.99)	0.112 (0.65)	0.027 (0.63)	1601	0.0%
6	0.211 (0.57)	0.075 (0.59)	0.041 (0.46)	2209	0.1%
7	0.116 (0.70)	0.054 (0.65)	0.066 (0.18)	2891	0.2%

To check the last possibility, I divide average order imbalance size by deciles and calculate average earnings surprise variable. The average order imbalance is calculated from order imbalance data between 7 days before an announcement and 2 days before the announcement. Note that earnings surprise variable has range of -0.5 ~ 0.5, and 0.1 is the difference between two close earnings surprise deciles. Table 3 shows that earnings surprise variable does not have a significant pattern. Non-monotonic relation shows that the rank of order imbalance is not a good method to forecast an upcoming announcement. The difference between two extreme order imbalance deciles (Min and Max) is the largest for order imbalance of trades, but it is only 0.12. Using extreme values of order imbalance can explain just one rank difference between two close earnings surprise deciles.

The simplest form of model specification would be using only signs of order imbalance (positive vs. negative). In table 4, difference between positive order imbalance and negative order imbalance is 0.05, which means the sign of order imbalance can explain less than one rank change of earning surprise deciles.

One may argue that one day's order imbalance is not an appropriate measure of informed trading. Kyle (1985) shows informed investors will split her orders lest her trading is detected by public. So I aggregate several days of order imbalance and check whether the aggregation can increase the predictive power of order imbalance. In table 5, I take 5-day moving average of order imbalance and test the relation between the moving average and earnings surprise.

$$v_{it} = \alpha + \delta_1 \cdot (\sum_{i=j-4}^{j} OI_{it-j}) \cdot (\lambda_{it-j}) + \delta_2 \cdot r_{i,t-j-1} + \varepsilon_{it} \qquad (5)''$$

Table 3 – Order imbalance size and earnings surprise

For each stock, I take average order imbalance between 7 days before announcement and 2 days before announcement. Then I rank the average order imbalance into deciles. Following table shows average earnings surprise by the deciles. Earnings surprise takes values from -0.5 to 0.5.

Panel A: Rank by order imbalance of shares			
Order imbalance rank	*Earnings surprise (Average)*	*Earnings surprise (Std. deviation)*	*Observations*
Low (most negative)	-0.048	0.321	291
2	-0.034	0.303	293
3	-0.027	0.312	297
4	-0.022	0.322	291
5	0.009	0.320	291
6	0.001	0.309	288
7	0.052	0.307	301
8	0.014	0.317	286
9	0.031	0.317	295
High (most positive)	0.004	0.326	292

Panel B: Rank by order imbalance of dollars			
Order imbalance rank	*Earnings surprise (Average)*	*Earnings surprise (Std. deviation)*	*Observations*
Low (most negative)	-0.051	0.319	295
2	-0.034	0.308	289
3	-0.027	0.309	295
4	-0.020	0.320	287
5	0.006	0.322	299
6	0.002	0.306	284
7	0.052	0.312	299
8	0.020	0.316	295
9	0.026	0.317	290
High (most positive)	0.004	0.326	292

Panel C: Rank by order imbalance of trades			
Order imbalance rank	*Earnings surprise (Average)*	*Earnings surprise (Std. deviation)*	*Observations*
Low (most negative)	-0.084	0.299	294
2	-0.028	0.298	296
3	0.014	0.316	290
4	0.017	0.327	288
5	-0.029	0.308	297
6	-0.019	0.333	290
7	0.024	0.313	289
8	0.028	0.319	294
9	0.019	0.321	295
High (most positive)	0.042	0.324	292

Table 4 – Positive order imbalance vs. negative order imbalance

For each stock, I take average order imbalance between 7 days before announcement and 2 days before announcement. Then I divide the average order imbalances by their signs. Following table shows average earnings surprise by order imbalance signs. Earnings surprise takes values from -0.5 to 0.5.

Panel A: Order imbalance of shares

Order imbalance sign	*Earnings surprise (Average)*	*Earnings surprise (Std. deviation)*	*Observations*
Positive	0.020	0.315	1664
Negative	-0.031	0.316	1261

Panel B: Order imbalance of dollars

Order imbalance sign	*Earnings surprise (Average)*	*Earnings surprise (Std. deviation)*	*Observations*
Positive	0.021	0.315	1695
Negative	-0.033	0.316	1230

Panel C: Order imbalance of trades

Order imbalance sign	*Earnings surprise (Average)*	*Earnings surprise (Std. deviation)*	*Observations*
Positive	0.013	0.320	1688
Negative	-0.022	0.310	1237

Table 5 – Moving average of order imbalance and private information

$$v_{it} = \alpha + \delta_1 \cdot (\sum_{i=j-4}^{j} OI_{it-j}) \cdot (\lambda_{it-j}) + \delta_2 \cdot r_{i,t-j-1} + \varepsilon_{it} \qquad (5)''$$

Here I use moving average of past 5 days' order imbalance. Earnings surprise measure v is derived from the difference between actual earnings and average analyst forecasts. I rank the earnings surprise into 11 ranks, divide it by 10, and subtract 0.5. *OI* is moving average of past 5 days' order imbalance and r_{t-1} is previous day's mid-quote stock return. For every trading day, I use 250 prior business days of order imbalance data to estimate Kyle's lambda. I regress stock return on order imbalance, and Kyle's lambda is the coefficient. Regression method for equation (3) is OLS with clustering and heteroskedasticity controlled error structure. P-values are in the parenthesis. Coefficients significant in 1%, 5%, and 10% level are marked with a, b, and c.

Panel A: Regression with order imbalance of shares				
Days before announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	1.414[b] (0.04)	0.222 (0.42)	1578	0.4%
3	0.598 (0.35)	0.289[c] (0.07)	1629	0.2%
4	-0.358 (0.58)	0.188 (0.20)	1924	0.1%
5	0.419 (0.63)	0.084 (0.76)	1601	0.0%
6	1.278[c] (0.08)	0.044 (0.78)	2209	0.2%
7	1.114[c] (0.06)	0.051 (0.70)	2891	0.1%

Panel B: Regression with order imbalance of dollars				
Days before announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	1.432[b] (0.04)	0.222 (0.42)	1578	0.4%
3	0.583 (0.37)	0.289[c] (0.07)	1629	0.2%
4	-0.382 (0.56)	0.189 (0.20)	1924	0.1%
5	0.388 (0.66)	0.086 (0.76)	1601	0.0%
6	1.269[c] (0.09)	0.044 (0.78)	2209	0.2%
7	1.100[c] (0.06)	0.052 (0.69)	2891	0.1%

Panel C: Regression with order imbalance of trades				
Days before announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	1.731[a] (0.00)	0.175 (0.51)	1578	0.8%
3	1.158[b] (0.01)	0.211 (0.19)	1629	0.5%
4	0.275 (0.56)	0.172 (0.24)	1924	0.1%
5	0.593 (0.34)	0.053 (0.85)	1601	0.1%
6	1.171[b] (0.02)	0.021 (0.89)	2209	0.3%
7	1.063[a] (0.01)	0.028 (0.83)	2891	0.3%

Table 5 shows moving average does not significantly improve the predictive power of order imbalance. The average of order imbalance gains some significance at day 6 and 7, though. Order imbalance of trades has the highest significance, while other measures show significance around 10% level. So an investor who wants to use order imbalance as a predictor should focus on order imbalance of trades and aggregate several days of data. The size of coefficients is still disappointing. Significant coefficients are around 1.00. Since Kyle's lambda measure have values around 0.12, this result indicates that a completely buy-side skewed order imbalance of 1 predicts at most a 0.1 rank higher earnings surprise.

Overall, I find weak evidence that order imbalance predicts forthcoming information. Order imbalance may be too noisy to be used as an indicator of forthcoming information.

B. Order imbalance and public information

According to hypothesis 2, if US stock market is efficient, quote change should absorb the effect of earnings announcements before any trade takes in place. Then order imbalance should be independent of earnings surprise.

H2: Order imbalance is independent of public announcement.

To test hypothesis 2, I directly measure the correlation between order imbalance and earnings surprise using multiple regression. I take 3-day window around earnings announcement and check the relation between order imbalance and earnings surprise during this period. Kothari and Warner (2006) show that the power of short term event studies has little dependence to how researcher specifies an event return. Using raw order imbalance in regression would not problematic.[18] To compare with the result in the previous section, I use the same equation (5) with order imbalances after announcement dates.

$$v_{it} = \alpha + \delta_1 \cdot (OI_{it+j}) \cdot (\lambda_{it+j}) + \delta_2 \cdot r_{i,t+j-1} + \varepsilon_{it} \qquad (5)$$

Table 6 has the result for equation (5). The result from order imbalance of shares or dollars shows that the order imbalance in general moves independent of the public announcement. This result implies that market makers are fast to adjust their quotes. Meanwhile, order imbalance of trades has some positive correlation with earnings announcement.

I use a different model specification to further investigate the relation between public information and order imbalance. There is a possibility that the reaction of order imbalance around earnings announcements is correlated with other variables such as firm characteristics or analyst coverage. Lo and MacKinley (1990) and Brennan, Jegadeesh, and Swaminathan (1993) show larger firm size and higher analyst coverage is related to the faster speed of price discovery. Barber, Lehavy, McNicholes, Trueman (2001) find analysts' EPS forecasts are positively skewed, and so stocks with many forecasts would have more difficulty to meet analyst forecasts. So I introduce equation (6), which uses order imbalance as dependent variable and includes market value and analyst coverage as control variables.

---

[18] One can control the market-wide effect or industry effect by subtracting the average order imbalances. Market or industry effect controlled order imbalances yield similar results to raw order imbalance. It is probably because the average of multiple firms' order imbalances often approaches to zero.

Table 6 – Order imbalance and public information

$$v_{it} = \alpha + \delta_1 \cdot (OI_{it-j}) \cdot (\lambda_{it-j}) + \delta_2 \cdot r_{i,t-j-1} + \varepsilon_{it} \qquad (5)$$

Earnings surprise measure *v* is derived from the difference between actual earnings and average analyst forecasts. I rank the earnings surprise into 11 ranks, divide it by 10, and subtract 0.5. *OI* is order imbalance and r_{t-1} is previous day's mid-quote stock return. For every trading day, I use 250 prior business days of order imbalance data to estimate Kyle's lambda. I regress stock return on order imbalance, and Kyle's lambda is the coefficient. Regression method for equation (3) is OLS with clustering and heteroskedasticity controlled error structure. P-values are in the parenthesis. Coefficients significant in 1%, 5%, and 10% level are marked with a, b, and c.

Panel A: Regression with order imbalance of shares

Days around announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
Previous day	-0.394 (0.32)	0.079 (0.63)	2408	0.1%
Announcement day	0.328 (0.36)	0.436[b] (0.01)	2925	0.3%
Next day	0.818[c] (0.07)	0.054 (0.79)	2258	0.2%

Panel B: Regression with order imbalance of dollars

Days around announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
Previous day	0.874 (0.16)	0.070 (0.66)	2408	0.1%
Announcement day	0.479 (0.38)	0.442[a] (0.00)	2925	0.3%
Next day	1.150 (0.11)	0.032 (0.88)	2258	0.1%

Panel C: Regression with order imbalance of trades

Days around announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
Previous day	0.546[b] (0.03)	0.076 (0.63)	2408	0.2%
Announcement day	0.415[c] (0.07)	0.437[a] (0.00)	2925	0.4%
Next day	0.824[a] (0.00)	0.080 (0.70)	2258	0.3%

$$OI_{i,q} = \alpha + \gamma_1 \cdot surprise_{i,q} + \gamma_2 \cdot mkt\_value_{i,q} + \gamma_3 \cdot num\_analysts_{i,q} + \varepsilon_{i,q} \qquad (6)$$

'*OI*' stands for order imbalance around the earnings announcement and '*surprise*' stands for the earnings surprise. '*OI*' is calculated by aggregating 3 day's order imbalances – a day before, announcement date, and a day after. 'Mkt_value' is the monthly average market value of the firm, and it is measured 2 months before the announcement date. This setting is to prevent quarterly earnings announcements from interfering with the market value. 'Num_analysts' is the number of analysts who made EPS forecasts. The minimum of this variable is 5. Because I estimate equation (6) in three day period around earnings announcement, date dummies – indicating whether the date is before, on, or after announcement date – are in equation (6) to control the fixed effect. Further, to see whether stock liquidity matters, I report volume stratified result. The volume is average monthly volume 2 months before the announcement, and in each year, I rank the volume into quartiles. The estimation method is again heteroskedasticity and clustering controlled OLS.

Table 7 shows that earnings surprise does not have positive and significant correlation with order imbalance. Order imbalance of trades seem to have a positive and significant correlation but it is restricted to low trading volume stocks. This result implies that in general the US stock market exhibits rapid quote revisions such that the market's price discovery process is efficient. Quotes can offset the effect of the announcement without much help from trading.

Table 7 – Order imbalance around earnings announcements

$$OI_{i,q} = \alpha + \gamma_1 \cdot surprise_{i,q} + \gamma_2 \cdot mkt\_value_{i,q} + \gamma_3 \cdot num\_analysts_{i,q} + \varepsilon_{i,q} \qquad (6)$$

OI stands for order imbalance of stock *i* at quarter *q*. *Surprise* is the ranked earnings surprise of stock *i* at quarter *q*. *Mkt_value* is the monthly average of market value, and it comes from 2 months before the announcement. *Num_analysts* counts the number of analyst forecasts. Dummy variables are included to control the date effect – whether the date is before, on, or after earnings announcement date. In volume stratified result, I rank monthly average volume into quartiles. The monthly average volume is from 2 months before an earnings announcement. I use OLS with heteroskedasticity corrected errors accounting for clustering by stock or month. P-values are in the parenthesis. All the coefficients are multiplied by 10^3. Coefficients significant in 1%, 5%, and 10% level are marked with a, b, and c.

Panel A: Order imbalance of shares

	All stocks	Volume Stratified Result			
		Low volume stocks	Mid-low volume stocks	Mid-high volume stocks	High volume stocks
Surprise rank	6.008 (0.40)	22.273 (0.28)	-1.982 (0.89)	-2.397 (0.83)	5.314 (0.49)
Market value	0.036 (0.24)	11.807 (0.12)	0.400 (0.89)	0.531 (0.51)	0.025 (0.36)
Number of Forecasts	0.828[b] (0.02)	-1.780 (0.60)	1.505 (0.32)	0.582 (0.55)	-0.068 (0.83)
Observations	2923	1892	1900	1898	1895
Adj. R-square	0.1%	0.2%	0.1%	0.1%	0.3%

Panel B: Order imbalance of dollars

	All stocks	Volume Stratified Result			
		Low volume stocks	Mid-low volume stocks	Mid-high volume stocks	High volume stocks
Surprise rank	5.703 (0.43)	22.155 (0.29)	-2.162 (0.88)	-2.849 (0.80)	4.953 (0.52)
Market value	0.033 (0.28)	11.337 (0.13)	0.258 (0.93)	0.470 (0.56)	0.022 (0.40)
Number of Forecasts	0.801[b] (0.02)	-1.763 (0.60)	1.484 (0.32)	0.557 (0.56)	-0.067 (0.83)
Observations	2923	1892	1900	1898	1895
Adj. R-square	0.1%	0.2%	0.1%	0.1%	0.3%

Panel C: Order imbalance of trades

	All stocks	Volume Stratified Result			
		Low volume stocks	Mid-low volume stocks	Mid-high volume stocks	High volume stocks
Surprise rank	18.635[c] (0.07)	57.231[a] (0.01)	10.700 (0.46)	20.158 (0.11)	-17.035 (0.58)
Market value	0.082[b] (0.05)	11.286[c] (0.09)	0.394 (0.89)	2.248[b] (0.02)	0.094[b] (0.02)
Number of Forecasts	1.276[a] (0.01)	2.166 (0.49)	0.606 (0.67)	1.221 (0.22)	-0.892 (0.43)
Observations	2923	1892	1900	1898	1895
Adj. R-square	0.2%	0.8%	0.1%	0.7%	0.2%

C. Order imbalance and past information

The results in previous two sections do not violate efficiency market hypothesis. For the private information case, informed investors may be effective in hiding their trades. For the public information case, order imbalance should be independent of the information lest anyone benefits from public information. In a similar vein, past information should have no effect on current trades. As we see in the case of public information, quote changes absorb the effect of the announced information. Even if earnings announcements are serially correlated (see Chordia and Shivakumar 2006, for example), earnings surprise would not be, because financial analysts will account for the serial correlation and adjust their earnings forecasts accordingly.

H3: Order imbalance is independent of past announcement.

I use equation (5) again to test hypothesis 3. This setting makes the results comparable. This time, order imbalance data is order imbalance occurring after earnings announcements. The estimation method is OLS with a heteroskedasticity robust and clustering corrected error structure. I control for clustering by firm, year, and quarter.

$$v_{it} = \alpha + \delta_1 \cdot (OI_{it-j}) \cdot (\lambda_{it-j}) + \delta_2 \cdot r_{i,t-j-1} + \varepsilon_{it} \qquad (5)$$

Table 8 contains the regression results. Order imbalance starts to reflect past earnings surprises after the announcement. There are more days with positive and significant coefficients. Note that I did not change the model specification at all compared to the previous tests. The positive correlation between order imbalance and past information challenges the assumption that order imbalance captures informed trading. If the market is efficient, investors would not trade based on past information. Table 9 shows the result with a moving average. The relation between past announcements and order imbalance gets stronger. Compared to the test before earnings announcements (table 5), not only there are more days with significant coefficients but also the coefficients are larger. This result suggests that investors consistently trade based on past earnings surprises.

Of course, there can be some hidden private information after earnings announcements. If so, the stock return will gradually adjust to the information, and order imbalance may be predicting the information. To check this possibility, I test the correlation between order imbalance after announcement and future stock return. Indeed, the accounting literature finds a market anomaly that suggests this conjecture. Stocks with high (low) earnings surprises continue to have high (low) stock returns until the next earnings announcement date. This phenomenon is called as post earnings announcement drift (PEAD). Recent papers find several market microstructure variables are related in PEAD. Mendenhall (2004) shows PEAD is related to bid-ask spread, Sadka (2006) argues stock liquidity plays important role to PEAD, and Garfinkel and Sokobin (2006) finds abnormal trading volume around announcement dates is linked to PEAD. Order imbalance could be another variable related to PEAD.

Table 8 – Order imbalance after earnings announcements

$$v_{it} = \alpha + \delta_1 \cdot (OI_{it+j}) \cdot (\lambda_{it+j}) + \delta_2 \cdot r_{i,t+j-1} + \varepsilon_{it} \qquad (5)$$

Earnings surprise measure *v* is derived from the difference between actual earnings and average analyst forecasts. I rank the earnings surprise into 11 ranks, divide it by 10, and subtract 0.5. *OI* is order imbalance and r_{t-1} is previous day's mid-quote stock return. For every trading day, I use 250 prior business days of order imbalance data to estimate Kyle's lambda. I regress stock return on order imbalance, and Kyle's lambda is the coefficient. Regression method for equation (3) is OLS with clustering and heteroskedasticity controlled error structure. P-values are in the parenthesis. Coefficients significant in 1%, 5%, and 10% level are marked with a, b, and c.

Panel A: Regression with order imbalance of shares

Days after announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	0.680 (0.16)	0.633[a] (0.00)	1614	0.7%
3	0.834[b] (0.02)	0.604[a] (0.00)	1678	0.9%
4	0.838[b] (0.05)	0.119 (0.46)	1200	0.4%
5	0.766 (0.12)	0.172 (0.23)	1244	0.4%
6	0.538 (0.16)	0.075 (0.46)	2350	0.1%
7	0.856[b] (0.01)	0.423[a] (0.00)	2827	0.7%

Panel B: Regression with order imbalance of dollars				
Days after announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	0.678 (0.17)	0.634[a] (0.00)	1614	0.7%
3	0.824[b] (0.02)	0.605[a] (0.00)	1678	0.9%
4	0.843[b] (0.04)	0.120 (0.45)	1200	0.4%
5	0.761 (0.12)	0.175 (0.23)	1244	0.4%
6	0.530 (0.17)	0.075 (0.46)	2350	0.1%
7	0.833[b] (0.01)	0.422[a] (0.00)	2827	0.7%

Panel C: Regression with order imbalance of trades				
Days after announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	0.613[c] (0.09)	0.638[a] (0.00)	1614	0.8%
3	0.230 (0.39)	0.659[a] (0.00)	1678	0.7%
4	0.557[c] (0.05)	0.141 (0.39)	1200	0.4%
5	0.655[c] (0.05)	0.200 (0.17)	1244	0.5%
6	0.388[c] (0.10)	0.089 (0.38)	2350	0.1%
7	0.809[a] (0.00)	0.449[a] (0.00)	2827	0.9%

Table 9 – Moving average of order imbalance after earnings announcements

$$v_{it} = \alpha + \delta_1 \cdot (\sum_{i=j-4}^{j} OI_{it-j}) \cdot (\lambda_{it-j}) + \delta_2 \cdot r_{i,t-j-1} + \varepsilon_{it} \qquad (5)''$$

Earnings surprise measure *v* is derived from the difference between actual earnings and average analyst forecasts. I rank the earnings surprise into 11 ranks, divide it by 10, and subtract 0.5. *OI* is order imbalance and r_{t-1} is previous day's mid-quote stock return. For every trading day, I use 250 prior business days of order imbalance data to estimate Kyle's lambda. I regress stock return on order imbalance, and Kyle's lambda is the coefficient. Regression method for equation (3) is OLS with clustering and heteroskedasticity controlled error structure. P-values are in the parenthesis. Coefficients significant in 1%, 5%, and 10% level are marked with a, b, and c.

Panel A: Regression with order imbalance of shares

Days after announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	0.368 (0.65)	0.646[a] (0.00)	1614	0.6%
3	0.735 (0.29)	0.634[a] (0.00)	1678	0.7%
4	2.230[a] (0.00)	0.106 (0.47)	1200	0.9%
5	1.567[b] (0.03)	0.122 (0.42)	1244	0.5%
6	1.681[a] (0.01)	0.032 (0.76)	2350	0.3%
7	1.978[a] (0.00)	0.392[a] (0.00)	2827	0.9%

Panel B: Regression with order imbalance of dollars

Days after announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	0.357 (0.67)	0.647[a] (0.00)	1614	0.6%
3	0.723 (0.30)	0.636[a] (0.00)	1678	0.7%
4	2.240[a] (0.00)	0.106 (0.47)	1200	0.9%
5	1.579[b] (0.03)	0.123 (0.41)	1244	0.5%
6	1.684[a] (0.01)	0.034 (0.74)	2350	0.3%
7	1.965[a] (0.00)	0.391[a] (0.02)	2827	0.9%

Panel C: Regression with order imbalance of trades

Days after announcement	δ_1	δ_2	*Observations*	*Adj. R-square*
2	1.162[b] (0.05)	0.599[a] (0.00)	1614	0.9%
3	0.901[c] (0.09)	0.596[a] (0.00)	1678	0.8%
4	1.636[a] (0.00)	0.106 (0.46)	1200	0.9%
5	1.655[a] (0.00)	0.090 (0.55)	1244	1.0%
6	1.636[a] (0.00)	0.010 (0.92)	2350	0.6%
7	1.645[a] (0.00)	0.384[a] (0.00)	2827	1.1%

The construction of post-earnings-announcement-drift variable follows Mendenhall (2004). He measures the drift by subtracting size deciles portfolio returns from cumulative stock returns between two earnings announcements dates. Since I use order imbalances up to 7 business days after the announcement, I take out overlapping days when calculating the PEAD. I drop an observation if the period between announcements is more than 120 calendar days, so that I only measure the return between quarterly announcements. Cumulative abnormal returns (CAR) after earnings announcements are therefore:

$$CAR_{i,q} = \Pi_{T_q+14}^{T_{q+1}-1}(1+r_i) - \Pi_{T_q+14}^{T_{q+1}-1}(1+EWR_i)$$

where Tq indicates the date of an earnings announcement at quarter q, r is for return of stock i, and EWR means equal weighted portfolio return of same size deciles in the CRSP file

My test of the relation between order imbalance and post-earnings-announcement-drift (PEAD) is as follows:

$$CAR_{i,q} = \alpha + \phi_t \cdot OI_{i,t} + \theta \cdot surprise_{i,q} + \varepsilon_{i,q} \qquad (7)$$

Equation (7) tests whether any of the order imbalances after earnings announcements is correlated with CAR. Small t stands for the days after announcement dates. Earnings surprise is included as a control variable, because PEAD was originally constructed to measure the correlation between CARs and earnings surprises.[19]

Table 10 shows order imbalance is not related to CAR. On the other hand, correlation between earnings surprise and CAR is robust. I reject the possibility that order imbalance

[19] Due to weekends and holidays, I cannot run the equation with more than 5 days of order imbalances.

at earnings announcements contains some forthcoming information. The order imbalance after the earnings announcement is therefore acquiring its momentum from past earnings surprise. Thus the order imbalance pattern following a major information event can merely be reflecting past information. This result is inconsistent with efficient market hypothesis.

Table 10 – Order imbalance and post-earnings-announcement-drift (PEAD)

$$CAR_{i,q} = \alpha + \phi_t \cdot OI_{i,t} + \theta \cdot surprise_{i,q} + \varepsilon_{i,q} \qquad (7)$$

CAR stands for cumulative abnormal return between two consecutive earnings announcements. *OI* is order imbalance after announcement and *surprise* is earnings surprise at the announcement. I use OLS with heteroskedasticity corrected errors accounting for clustering by stock or month. P-values are in the parenthesis. Coefficients significant in 1%, 5%, and 10% level are marked with a, b, and c.

Panel A: Order imbalance of shares					
	Model 1	Model 2	Model 3	Model 4	Model 5
Order Imbalance (announcement + 2 days)	-0.043 (0.26)			-0.069 (0.33)	
Order Imbalance (announcement + 3 days)	0.002 (0.95)			0.029 (0.67)	
Order Imbalance (announcement + 4 days)		-0.005 (0.95)		0.38 (0.49)	
Order Imbalance (announcement + 5 days)		-0.114 (0.21)			-0.110 (0.14)
Order Imbalance (announcement + 6 days)			0.072 (0.05)		0.072 (0.32)
Order Imbalance (announcement + 7 days)			0.007 (0.84)		0.031 (0.66)
Earnings surprise	0.061[a] (0.00)	0.144[a] (0.00)	0.083[a] (0.00)		
Observations	933	570	1858	251	760
Adj. R-square	1.0%	1.9%	1.0%	0.3%	0.4%

Panel B: Order imbalance of dollars					
	Model 1	Model 2	Model 3	Model 4	Model 5
Order Imbalance (announcement + 2 days)	-0.043 (0.27)			-0.061 (0.40)	
Order Imbalance (announcement + 3 days)	0.002 (0.96)			0.026 (0.71)	
Order Imbalance (announcement + 4 days)		-0.010 (0.90)		0.32 (0.57)	
Order Imbalance (announcement + 5 days)		-0.102 (0.25)			-0.100 (0.18)
Order Imbalance (announcement + 6 days)			0.073[b] (0.04)		0.071 (0.33)
Order Imbalance (announcement + 7 days)			0.008 (0.82)		0.033 (0.63)
Earnings surprise	0.061[a] (0.00)	0.144[a] (0.00)	0.083[a] (0.00)		
Observations	933	570	1858	251	760
Adj. R-square	1.0%	1.8%	1.1%	0.5%	0.3%

D. Order imbalance and stock return

Previous section suggests that some order imbalances may not be related to information. Market makers may be reluctant to change her quote level if order imbalances seem to be reflecting past information. Then order imbalances before earnings announcement can have smaller effect to stock returns compared to order imbalances after the announcement. On the other hand, it can be hard for market makers to differentiate order flows by their probability of informed trading. Market makers may take a simple approach and change their quote level according to all order imbalances. In this case, stock returns will be equally sensitive to any types of order imbalances.

I test whether order imbalances before earnings announcements have bigger effect on stock returns compared to order imbalances after earnings announcements. This test will show which type of order imbalance – information related or not – is more important to stock price change. Following Chordia and Subrahmanyam (2004), I include lagged order imbalances in the test equation:

$$r_{i,t} = \alpha + \sum_{j=0}^{5} \gamma_j \cdot OI_{i,t-j} + \varepsilon_{i,q} \qquad (8)$$

where r_{it} is daily mid-quote stock return of stock *i* at day *t* and OI_{it} is daily order imbalance of stock *i* at day *t-j*

Equation (8) tests the effect of contemporary and lagged order imbalances to stock return. I divide my dataset into two groups: stock returns before earnings announcements and stock returns after earnings announcements. The former dataset has stock returns between [-8, -1] days of earnings announcements and the latter dataset has stock returns between [6, 13] days of earnings announcements. The latter dataset has returns past 6 business days of earnings announcements to prevent using order imbalances before earnings announcements in the estimation.

Table 11 shows two types of order imbalance have similar effect to stock returns. Order imbalances after earnings announcements have slightly higher R-square and larger coefficients. This result indicates that non-information based order flows can equally move stock prices as information based order flows. Such effect can make price discovery processes more difficult and time consuming.

Table 11 – Order imbalance and stock return

$$r_{i,t} = \alpha + \sum_{j=0}^{5} \gamma_j \cdot OI_{i,t-j} + \varepsilon_{i,q} \qquad (8)$$

r_{it} is daily mid-quote stock return of stock *i* at day *t* and OI_{it} is daily order imbalance of stock *i* at day *t-j*. I divide my dataset into two groups: stock returns before earnings announcements and stock returns after earnings announcements. The former dataset has stock returns between [-8, -1] days of announcements and the latter dataset has stock returns between [6, 13] days of announcements. I use OLS with heteroskedasticity corrected errors accounting for clustering by stock or month. P-values are in the parenthesis. Coefficients significant in 1%, 5%, and 10% level are marked with a, b, and c.

Panel A: Order imbalances before earnings announcements			
	Order imbalance of shares	Order imbalance of dollars	Order imbalance of trades
Order Imbalance$_t$	0.044[a] (0.00)	0.044[a] (0.00)	0.052[a] (0.00)
Order Imbalance$_{t-1}$	0.015[a] (0.00)	0.015[a] (0.00)	0.026[a] (0.00)
Order Imbalance$_{t-2}$	-0.012[a] (0.00)	-0.012[a] (0.00)	-0.016[a] (0.00)
Order Imbalance$_{t-3}$	-0.008[a] (0.00)	-0.007[a] (0.00)	-0.009[a] (0.00)
Order Imbalance$_{t-4}$	-0.005[b] (0.01)	-0.005[b] (0.01)	-0.004[c] (0.06)
Order Imbalance$_{t-5}$	-0.005[b] (0.01)	-0.005[b] (0.01)	-0.010[a] (0.00)
Observations	11797	11797	11797
Adj. R-square	4.4%	4.4%	8.2%

Panel B: Order imbalances after earnings announcements			
	Order imbalance of shares	Order imbalance of dollars	Order imbalance of trades
Order Imbalance$_t$	0.052[a] (0.00)	0.052[a] (0.00)	0.054[a] (0.00)
Order Imbalance$_{t-1}$	0.022[a] (0.00)	0.022[a] (0.00)	0.028[a] (0.00)
Order Imbalance$_{t-2}$	-0.012[a] (0.00)	-0.012[a] (0.00)	-0.018[a] (0.00)
Order Imbalance$_{t-3}$	-0.002 (0.30)	-0.002 (0.29)	-0.001 (0.85)
Order Imbalance$_{t-4}$	-0.004[b] (0.03)	-0.004[b] (0.03)	0.002 (0.82)
Order Imbalance$_{t-5}$	-0.002 (0.38)	-0.002 (0.39)	0.001 (0.80)
Observations	11439	11439	11439
Adj. R-square	4.8%	4.8%	8.3%

V. Conclusion

Order imbalance may signal underlying information, and it can be a powerful supplement to the signals from stock returns or trading volume. However, its relation with information has not been thoroughly investigated. In order to use order imbalance in future applications, there should be more studies of its characteristics.

I empirically test the relation between order imbalance and information using earnings announcements as my information source. Before an earnings announcement, order imbalance has poor predictive power for the subsequently released earnings announcements. There are two explanations for this result; one is that informed investors are successful in hiding their trades and the other is that order imbalance moves independently of information. The former explanation is consistent with weak form market efficiency hypothesis that past data cannot predict an upcoming event.

When information is publicly announced, the order imbalance at announcement period moves independently of the released information. This independence is consistent with semi-strong form market efficiency hypothesis, that no investor should be able to make a profit from public information. The finding of independence is evidence that market makers quickly change their quotes to reflect the information before trades take in place.

Meanwhile, order imbalances after earnings announcements have a positive correlation with past earnings surprises. It is a puzzle why order imbalance is independent of information arriving at the announcement, but gains positive correlation afterwards. While the efficient market hypothesis is supported by the two earlier results, the relation between past information and order imbalance is inconsistent with efficient market hypothesis. I verify that order imbalances after announcement does not contain other information, but related to past earnings surprises. However, order imbalances after earnings announcements have similar effect to stock returns compared to that of order imbalances before announcements. This result indicates that non-information based order imbalances can have a significant effect to stock returns.

Overall, my results do not support the assumption that order imbalance is connected to informed trading. A framework that explains all three behavior patterns in order imbalance in this paper is yet to be developed. Thus, this evidence calls for additional research on the determinants of order imbalance, a variable that has potential to aid the investment decisions of many investors.

References

Amihud, Yakov and Mendelson, Heim, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics 17*, 223-249.

Andrade, Sandro, Chang, Charles, and Seasholes, Mark S., 2008, Trading imbalances, predictable reversals, and cross-stock effects, *Journal of Financial Economics 88*, 406-423.

Back, Kerry and Baruch, Schmuel, Information in securities markets: Kyle meets Glosten and Milgrom, *Econometrica 72*, 433-465.

Ball, Ray and Brown, Philip, 1968, An empirical evaluation of accounting income numbers, *Journal of Accounting Research 6*, 159-178.

Barber, Brad, Lehavy, Reuven, McNichols, Maureen, and Trueman, Brett, 2001, Can investors profit from the prophets? Security analyst recommendations and stock returns, *Journal of Finance 56*, 531-563.

Bartov, Eli., Radhakrishnan, Suresh, and Krinsky, Itzhak, 2000, Investor sophistication and patterns in stock returns after earnings announcements, *Accounting Review 75*, 43–63.

Bernard, Victor L. and Thomas, Jacob K., 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics 13*, 305-340.

Bhushan, R. 1994, An informational efficiency perspective on the post-earnings announcement drift, *Journal of Accounting and Economics 18*, 45–65.

Bollen, Nocholas P. and Whaley, Robert E., Does net buying pressure affect the shape of implied volatility functions?, *Journal of Finance 59*, 711-753.

Brandt, Michael. W. and Kavajecz, Kenneth. A., 2004, Price discovery in the U.S. treasury market: The impact of orderflow and liquidity on the yield curve, *Journal of Finance 59*, 2623-2654.

Brennan, Michael J., Jegadeesh, Narasimhan, and Swaminathan, Bhaskaran, 1993, Investment analysis and the adjustment of stock prices to common information, *Review of Financial Studies 6*, 799-824.

Brennan, Michael J., Chordia, Tarun, and Subrahmanyam, Avanidhar, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics 49*, 345-373.

Copeland, T. and Galai, D., 1983, Information effects on the bid-Ask spread, *Journal of Finance 5*, 1453-1469.

Chordia, Tarun, Roll, Richard, and Subrahmanyam, Avanidhar, 2008, Liquidity and market efficiency, *Journal of Financial Economics 87*, 249-268.

Chordia, Tarun and Shivakumar, Lakshmanan, 2006, Earnings and price momentum, *Journal of Financial Economics 80*, 627-656.

Chordia, Tarun and Subrahmanyam, Avanidhar, 2004, Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Economics 72*, 485-518.

Chordia, Tarun, Roll, Richard, and Subrahmanyam, Avanidhar, 2002, Order imbalance, liquidity and market returns, *Journal of Financial Economics 65*, 111-130.

Doyle, Jefferey T., Lundholm, Russel J., Mark T., 2007, The extreme future stock returns following I/B/E/S earnings surprises, *Journal of Accounting Research 44*, 849-887.

Easley, David, Engle, Robert F., O'Hara, Maureen, and Wu, Liuren, 2008, Time-varying arrival rates of informed and uninformed trades, *Journal of Financial Econometrics 6,* No. 2, pp. 171-208.

Easley, David, Kiefer, Nicholas M., and O'Hara, Maureen, 1997, One day in the life of a very common stock, *Review of Financial Studies* 10, No. 3, pp. 805-835.

Easley, David, Kiefer, Nicholas M., O'Hara, Maureen, Paperman, Joseph B., 1996, Liquidity, Information, and Infrequently Traded Stocks, *Journal of Finance 51*, No. 4, 1405-1436.

Easley, David and O'Hara, Maureen, 1987, Price, trade size and information in securities markets, *Journal of Financial Economics 19*, 69-90.

Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance*, *25*, 383-417.

Fleming, Michael J. and Remolona, Eli M., 1999, Price formation and liquidity in the U.S. treasury market: The response to public information, *Journal of Finance 54*, 1901-1915.

Garfinkel, Jon A., and Sokobin, Jonathan, 2006, Volume, opinion divergence and returns: A study of post-earnings-announcement-drift, *Journal of Accounting Research 44*, 85-112.

Glosten, Lawrence R. and Milgrom, Paul R., 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics 14*, 71-100.

Huang, Roger D. and Stoll, Hans R. Market microstructure and stock return predictions, *Review of Financial Studies 7*, 179-213.

Kandel, Eugene and Pearson, Neil D., Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy* 103, No. 4, 831-872.

Kothari, S.P. and Warner, Jerold B., 2006, Econometrics of event studies, Forthcoming in B. Espen Eckbo (ed.) *Handbook of corporate finance: empirical corporate finance*, volume A, Ch. 1.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica 53*, 1335-1355.

Lee, Charles M.C., 1992. Earnings news and small traders: an intraday analysis. *Journal of Accounting and Economics 15*, 265–302.

Lee, Charles M.C., and Ready, Mark J., 1992, Inferring trade direction from intraday data, *Journal of Finance 46*, 733-747.

Lo, Andrew, and MacKinlay, Craig, 1990, When are contrarian profits due to stock market overreaction?, *Review of Financial Studies 3*, 175-205.

Mendenhall, Richard R., 2004, Arbitrage risk and post-earnings-announcement drift, *Journal of Business 77*, 875-894.

Petersen, Mitchell A., 2007, Estimating standard errors in finance panel data sets: comparing approaches, *forthcoming Review of Financial Studies*.

Sadka, Ronnie, 2007, Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk, *Journal of Financial Economics 80*, 309-349.

Stoll, Hans R., 2000, Friction, *Journal of Finance* 55, 1479-1514.

Trueman, Brett, Wong, M. H. Franco and Zhang, Xiao-Jun, 2003, Anomalous stock returns around internet firms' earnings announcements, *Journal of Accounting and Economics 34*, 249-271.

Wooldridge, Jeffrey M., 2002, *Econometric analysis of cross section and panel data*, MIT Press, Cambridge, Massachusetts.

Appendix – The Construction of Order Imbalance Data

1. Criteria for stock selection are:

- Data source comes from Trade and Quote (TAQ) data.
- Data period is from January 1996 to December 2004.
- I exclude Certificates, ADRs, shares of beneficial interest, units, Americus Trust components, closed-end funds, preferred stocks and REITs from the dataset.
- I delete the stock is from the sample year if the price at any month-end during the year was greater than $999.
- I eliminate non-synchronous trading issue by marking stock return as missing if there was no trade on today or previous day.

2. When constructing order imbalance variable, I only use quotes and trades such that:

- Quotes and trades are in regular market trading times (from 9:30 to 16:00)
- There is no special settlement conditions
- All bid-ask spreads are positive

3. Method to calculate order imbalance is (Lee and Ready (1991) method):

- A trade is buyer (seller) initiated if it is closer to the ask (bid) of the prevailing quote.
- Prevailing quote should be at least 5 seconds old.

If the trade is at the midpoint of the quote, the trade is buyer (seller) initiated if prior stock price change was positive (negative).

CHAPTER 3

ORDER IMBALANCE AROUND SEASONED EQUITY OFFERINGS

I. Introduction

A seasoned equity offering (SEO) can create large order imbalances – buying pressure or selling pressure – around its issue date. An analysis of the trading pattern would be useful in understanding equity flotation cost, because the pattern shows how SEO characteristics affect stock trading activity in nearby periods. This paper studies stock order imbalance around an SEO and its relation to SEO underpricing.

There are many hypotheses on the observed trading pattern around SEOs, but most of them are tested using stock return data, rather than order imbalance information. Since stock price can change because of quote level changes as well as trading pressure, stock returns fail to give a clear picture of the role of trading pressure around SEOs. I use Lee and Ready (1992) method to directly estimate buying pressure / selling pressure around an SEO issue date and examine how order flow is related to major SEO characteristics including underpricing.

Analysis of order imbalances around security offerings is more common in the Initial Public Offering (IPO) literature than the SEO literature. (Ellis, Michaely, and O'Hara 2004, Aggarwal 2000, Boehmer and Fische 2003, and Lewellen 2006 for example) Many papers use proprietary data on institutional trading after an IPO. Ellis, Michaely, and O'Hara (2004) study market making activities after an IPO. The relation between price supports and trading patterns is investigated in Aggarwal (2000), Boehmer and Fische (2003), and Lewellen (2006). Meanwhile, the order imbalance of an SEO is first studied in Lease, Masulis, and Page (1991) and revisited by Huh and Subrahmanyam (2005). Cotter, Chen, and Kao (2004) analyze trading data to detect price stabilization activity around SEOs, while Autore (2007) and Henry and Koski (2008) study relation between short selling activity and stock price movements around an SEO. Order imbalance measure allows a researcher to capture abnormal trading patterns without using proprietary data, and so it can be useful to analyze SEOs. To my knowledge, this paper is the first attempt to analyze the relation between order imbalance and SEO characteristics, such as underpricing.

Indeed, order imbalance pattern is quite different from stock return pattern. SEO papers find negative stock returns before an issue date and positive returns afterwards. (see Kadlec, Loderer, and Sheehan 1994, Corwin 2003, and Meidan 2005 for example) Meanwhile, I find that order imbalance is slightly positive before an issue date and highly negative afterwards. This result is surprising because it is known that order imbalance and stock return moves to the same direction (Chordia and Subrahmanyam 2004). Huh and Subrahmanyam (2005) show that order imbalance and stock return shows negative correlation after an SEO issue date. Such inconsistency between trading pattern and stock return calls for in depth study on order imbalance around SEOs.

Based on the existing literature, I test five factors that may affect trading patterns around an SEO issue date. (1) Selling incentives generated by arbitrage seeking and supply shock, (2) Information asymmetry and uncertainty about SEO details, (3) Underwriter price support after post-SEO issuance, (4) Flipping activities[20], and (5) Market maker inventory management around SEOs. These factors give different predictions about order imbalance patterns and their effects on SEO underpricing. I study

[20] Flipping indicates acquiring the stocks in offering and selling them right after the issue date.

which factors drive trading activities around SEO issue dates, and how they are related to SEO underpricing and equity flotation costs more generally.

I find that positive order imbalances before an issue date can be explained by market maker inventory management. The positive order imbalance is generated by market makers lowering quotes to promote ask side transactions. More ask side transactions reduce market makers' inventory level before an SEO. As shown in Stoll (1978), the trading activities of market makers have little effect on stock price, because quotes should not be far away from current market price to promote bid side transactions.

If other types of investors engaged in selling activities before an SEO, there would be more bid side transactions, causing negative order imbalance. My test further confirms that the trading pattern is not affected by other SEO characteristics such as offer size or the downward slope of demand curve. This result implies that there are not excessive selling activities before SEO issue dates those which can push the price down.

Highly negative order imbalance after an issue date can be explained by underwriters' price support. Underwriters place limit orders at the bid to prevent the stock price from dropping below the offer price. As the price support gets stronger, there are more bid side transactions than ask side transactions, which will be measured as a negative order imbalance. I find that order imbalance following SEOs are negatively correlated with offer size and underwriter reputation. Also, order imbalances become more negative as the current market price drops below the offer price. These results indicate that price support activity is stronger when offer size is larger, underwriter reputation is stronger, and when the stock's market price drops below the offer price.

I find evidence that trading pressure can affect equity flotation cost. SEO underpricing is significantly correlated with order imbalance at the issue date. The correlation is significant, even after controlling for endogeneity and other known factors related to underpricing. This result implies that (1) trading pressure around an SEO can be an important factor that affects equity flotation cost, and (2) large underwriters can change the degree of underpricing by temporarily moving the stock's market price.

This paper makes two contributions to the literature. First, this paper shows that the trading patterns around an SEO are based on different economic mechanisms than that of stock returns. While price pressure and information asymmetry / uncertainty may be powerful hypotheses on stock return movement (see Corwin 2003), a stock's trading pattern is little affected by those factors. This paper shows that the trading patterns around an SEO are dominated by underwriters' price support activity and market makers' inventory adjustments. This evidence can be useful to SEO underwriters, traders, and regulators. Second, this paper shows that trading pressure can affect SEO underpricing. While underpricing is a major component of flotation costs, underwriters can reduce this issuer cost through their trading activity. Underwriters may even choose to short sell an SEO stock in order to impose a stronger price support (Henry and Koski 2008). In such cases, underpricing measures can convey a misleading signal to issuing firms about the success of its underwriter's advisory and selling activities.

The rest of paper is constructed as follows: Section 2 discusses hypotheses on trading pressure around an SEO. Section 3 describes order imbalance measure and SEO data. Section 4 shows how order imbalance is correlated with the factors of SEO. Section 5 studies relation between order imbalance and SEO underpricing. Section 6 concludes.

II. Hypotheses on Order imbalance around an SEO

A. Selling incentives

An SEO generates selling incentives to majority of traders. First, the offer price is usually lower than the market price. Gerard and Nanda (1993) show that this price difference creates an arbitrage opportunity for traders. Traders can short sell an SEO stock before an issue date and recover the position in primary market. Since this activity can push down the offer price further and discourage firms from issuing new shares, the SEC imposes restrictions on covering short selling position using newly offered shares.[21] Still, traders who can cover their short position elsewhere can short sell an SEO stock. Even underwriters can short sell before an issue date to control their exposure. Existing shareholders may also temporarily reduce their position in order to gain some short-term profit. There is still a ongoing debate on the effect of selling activity. Safieddine and Wilhelm (1996) find short selling restriction (SEC rule 10b-21) reduces underpricing, while Corwin (2003) and Kim and Shin (2004) argue the effect is marginal. Henry and Koski (2008) show short selling activity can cause a temporary price drop around an SEO issue date.

SEOs create another selling incentive, because they supply a large amount of shares to the market. The supply shock generates a temporary downward price pressure (see Kraus and Stoll 1972 for example), and traders may want to sell the stocks before they experience this downward price pressure effect. Some studies use stock return around an

[21] From 1988, SEC imposed Rule 10b-21, which prohibits covering short sales with shares from primary market. Rule 10b-21 applies to any short sales from announcement date to issue date. In 1997, SEC changed replaced Rule 10b-21 with Rule 105. Rule 105 prohibits short covering with shares from primary market, if short sales position is made within 5 days of issue date.

SEO as the evidence of price pressure. Kadlec, Loderer, and Sheehan (1994) and Meidan (2005) find stock returns are negative before the issue date and positive afterwards. Furthermore, Corwin (2003) shows SEO underpricing is increasing in offer size. Meanwhile, the stock return data of Altinkilic and Hansen (2003) does not fit the price pressure hypothesis. Scholes (1972) and Mikkelson and Partch (1985) find no correlation between offer size and price movements around the SEO issue date. They argue that in an efficient market, price pressure effect should be reflected in announcement dates rather than issue dates. In short, this remains a controversial area of research and motivates my analysis of trading patterns around an SEO issue date.

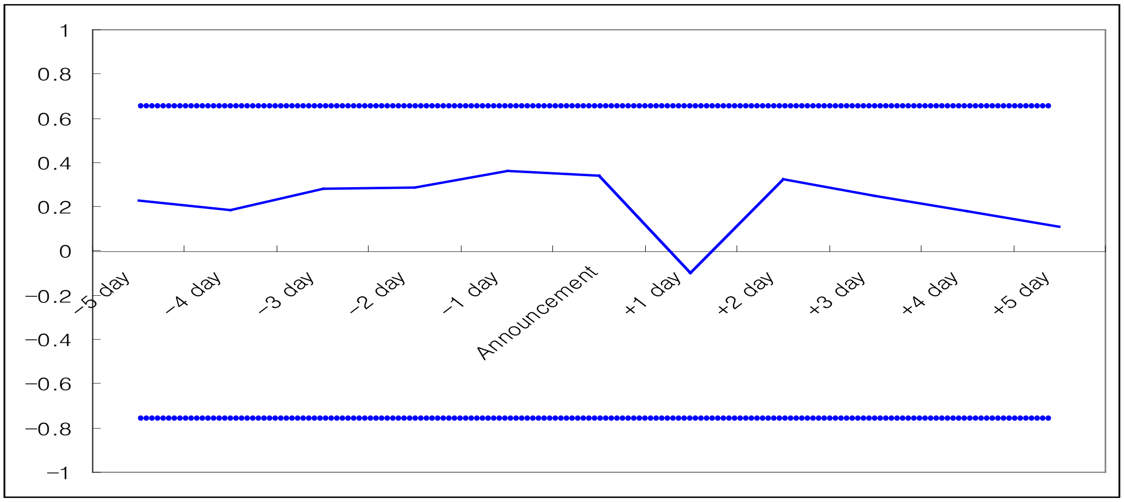
If trading patterns are dominated by the above mentioned selling incentives, order imbalances will show the following patterns. First, order imbalance should be negative before an issue date and positive afterwards. This prediction follows from the assumption that traders are selling the stock before an issue date, much like front running. Second, order imbalance before an issue date should be more negative as offer size increases. A larger offer size means more opportunities for short selling and greater price pressure. Third, order imbalance before and after an issue date should be similar in their size effects, but with opposite signs, because an SEO generates temporary selling incentives. Lastly, SEO underpricing should be negatively correlated with order imbalance. If those selling activities cause SEO underpricing, the degree of underpricing should be decreasing as order imbalance becomes less negative.

B. Uncertainty and information asymmetry

Since Myers and Majluf (1984), numerous papers confirmed that uncertainty and information asymmetry have an important bearing on the equity offering process. However, there can be two types of uncertainty and information asymmetry related to an SEO. Myers and Majluf (1984) theory focuses on manager's decision to issue equity. The effect should negatively affect stock price at an announcement date. I plot average order imbalance around SEO announcement dates in figure 1.

Figure 1 shows SEO announcement has insignificant effect to trading patterns. Average order imbalance is not highly negative around SEO announcement dates. Although SEO announcement may be a bad news, order imbalance may not be negative if quotes move fast enough to offset the announcement effect.

Since the actual offer size and offer price decisions are determined near the offer date, relatively un-informed investors can be reluctant to trade the stock before an issue date, given this increased uncertainty and information disadvantage.. In similar vein, Safieddine and Wilhelm (1996), Corwin (2003), Altinkinc and Hansen (2003) find uncertainty and information asymmetry affect SEO underpricing. They find that stocks with a higher degree of uncertainty and information asymmetry have a larger underpricing. Underpricing can be a result of setting conservative offer prices, but it also can be exacerbated by selling activities preceding the SEO issue date. Using order imbalance, we can test whether uncertainty and information asymmetry generates more selling activities before an SEO issue date.



- Average order imbalance higher or lower than dotted band is significant in 10% level.

Figure 1 – Order imbalance around an SEO announcement date

If uncertainty and information asymmetry before an issue date reduce investor demand for shares, then we would predict to observe negative order imbalance before SEO issue dates. The degree of uncertainty and information asymmetry should be negatively correlated with order imbalances prior to the SEO issue dates. If uncertainty and information asymmetry affects SEO underpricing through trading activities, then order imbalances should be negatively correlated with SEO underpricing. Uncertainty and information asymmetry before an issue date can reduce buying activities of traders. So this explanation predicts a negative order imbalance before an issue date. The degree of uncertainty and information asymmetry should be negatively correlated with order imbalance before an issue date. If uncertainty and information asymmetry affects SEO underpricing through trading activities, the order imbalance should be negatively correlated with SEO underpricing.

C. Underwriter price support

IPO studies such as Aggarwal (2000), Boehmer and Fishe (2003), and Lewellen (2006) find underwriters undertake considerable amounts of price support activity. Underwriters have the same incentive to support SEO stock prices. However, price support of SEO stocks is more complicated because of the typically larger trading volume in the secondary market post-SEO. It is also harder for a researcher to detect price support activity by studying stock return patterns, while order imbalances can more directly measure the degree of price support activity.

Price support activity is undertaken by placing large limit orders to buy stock. If underwriters place larger limit orders, then more traders can take advantage of this opportunity to sell shares at better prices. On the other hand, investors interested in buying shares would delay placing orders because the current quote levels are higher than the stock's intrinsic value due to the temporary underwriter price support activity. Thus, stronger price supports generate more buy side trading, resulting in negative order imbalances. Similarly, Cotter, Chen, and Kao (2004) argue that price stabilization trading occurs at bid. They also show that price stabilization gets stronger when market price falls below offer price.

In general, price support activity begins on the SEO issue date and can continue for a number of trading days thereafter. Therefore, price supports create negative order imbalances following SEO issue dates, while they have little known effect on order imbalances prior to the issue dates. Because the main purpose of price supports is to reduce underpricing, order imbalances should have a positive correlation with underpricing. More negative order imbalance would be a sign that underwriters are undertaking greater levels of price supports to reduce underpricing.

D. Flipping

Institutional investors who acquired shares in offering may sell the shares after an issue date. This type of sales is called as 'flipping'. Since offer price is typically lower than market price, institutional investors can sell their SEO shares in the secondary market at a higher price. This activity can give a short term profit. On the other hand, issuing firm and underwriters try to prevent flipping because it can signal sour views of future performance. Krigman, Shaw, and Womack (2001) find the degree of flipping activity is a predictor of future stock performance. Given the incentives for flipping, there is possibility that flipping generates a large selling pressure after an issue date. However, Chemmanur, He, and Hu (2007) argue flipping activities are rare in SEOs. They find about 3% of SEO shares are flipped.

I test whether flipping generates a significant selling pressure after an issue date. If flipping is a major determinant of trading activity around an SEO, we should observe following patterns: First, order imbalance should be negative after an issue date. Second, there will be positive correlation between order imbalance and underwriter reputation. Flipping harms post SEO stock performance and signals poor prospects, but better underwriters would reduce those effects. A higher underwriter reputation will push order imbalance toward positive side. Third, since flipping is profitable when secondary market price is higher than offer price, order imbalance would be more negative when current market price is higher than offer price. Lastly, order imbalance would be positively correlated with underpricing. Flipping creates negative order imbalance as well as lower stock return, so order imbalance and stock return should go in a same direction. Note that flipping hypothesis gives a similar prediction to underwriter price support hypothesis, except for the opposite effect of underwriter reputation and current market price.

E. Market maker inventory management

Lease, Masulis, and Page (1991) find that there are positive order imbalances before an SEO issue date. They explain this phenomenon by market maker inventory management. Since an SEO can induce large selling pressure in the secondary market as of the issue date, market maker inventory can significantly increase after an SEO. As in Stoll (1978), market makers have incentives to keep a constant inventory level, and so anticipating future selling pressure they would try to reduce their inventory levels before an SEO issue date. In order to reduce their inventory level, market makers purposely place their quote level below market price. If a quote level is below the share's usual price level, then buyers would want to trade, while sellers would stay away, generating a positive order imbalance. A positive order imbalance reduces market maker inventory because buyers can obtain stock from market makers. This type of trading can also yield short term profits for market makers. By changing quote levels in advance of SEO issue dates, market makers are effectively following a short selling strategy. They reduce inventory (sell shares to public) before an issue date and restore it to its ordinary level afterwards. The difference between ordinary selling and inventory management represents its effect on order imbalance. Ordinary selling creates a negative order imbalance, while inventory management generates a positive order imbalance.

If order imbalance follows market maker inventory management, there will be a positive order imbalance before an SEO issue date. A negative order imbalance may follow afterwards, but market makers also have option to buy shares in the primary market. The effect of order imbalance on underpricing can be marginal, because the order imbalance is generated by a temporary quote change rather than a permanent price change. As in Stoll (1978), market makers can reduce their inventory level only when price stays higher than quote level. Table 1 summarizes five factors and their predictions for order imbalance patterns. Because these factors provide different predictions, we can see which explanation fits the best by analyzing order imbalance pattern.

Table 1 – Hypotheses and predictions on order imbalance

This table shows predictions on order imbalance patterns around an SEO issue date.

Hypotheses:

1. Price pressure: The supply shock of an SEO creates a temporary downward selling pressure to a stock
2. Uncertainty and Information asymmetry: Uncertainty and information asymmetry of an SEO keep buyers from trading the stock before its issue date
3. Price support: Underwriters place large limit orders at bid after an issue date
4. Flipping: Institutional investors acquire shares in primary market and sell the shares in secondary market
5. Market maker inventory management: Market makers drop their quote level before an SEO issue date to reduce their inventory level

	Before an issue date	At an issue date	After an issue date	Overall change by an SEO	Correlation with underpricing
Selling incentives	Negative	Negative	Positive	Little change	Negative correlation
Uncertainty and Information asymmetry	Negative				Positive correlation
Underwriter price support		Negative	Negative	Negative	Positive correlation
Flipping			Negative	Negative	Positive correlation
Market maker inventory management	Positive	Negative	Negative	Little change	Little correlation

III. Data

A. Data

I use Trade and Quote (TAQ) data for ordinary common shares from 1996 to 2004 to construct my order imbalance dataset. The construction method is in appendix, and it closely follows the method of Chordia, Roll, and Subrahmanyam (2002). Their method is based on Lee and Ready (1991), but it imposes additional filters to reduce problems from scarce trading. The basic idea of the method is to count the number of shares traded at ask side (buy side orders) and bid side (sell side orders) every day. Subtracting shares traded at bid side from shares traded at ask side, I get a daily measure on order imbalance. For example, if there are more shares traded at ask side than bid side, there is a positive order imbalance at that day. Lee and Ready (1991), Chordia, Roll, and Subrahmanyam (2002), and Chordia and Subrahmanyam (2004) show this method gives an effective measure of daily price pressure.

I measure order imbalance in numbers of shares. If order imbalances are measured in number of trades or dollar volume, I find qualitatively similar results. I divide the share imbalance measure by total shares outstanding (excluding new shares to be issued) to normalized the price pressure measure. This metric facilitates a cross-sectional comparison across the sample of stocks with SEOs.

Seasoned equity offering data is extracted from SDC database. To be included in the dataset, a security offering must pass the following filters:

- The offer should be a public offer for a US common stock.
- Rights offerings and shelf offerings are excluded.

- Offer price should be more than $5 per share.
- Primary listing of the stock is NYSE, AMEX, or NASDAQ.
- Close-end funds/trusts, limited partnerships, LBO firms, firms with previous LBOs, private placements, unit investment trusts, unit issues, simultaneous offerings, and simultaneous international offerings are excluded.
- An issuer should be a US company.
- REITs and equity spinoffs are excluded.

Additionally, I use the CRSP and Market Microstructure databases available at Vanderbilt University to get stock price data and market microstructure data. The final dataset contains 1096 seasoned equity offerings.

Many papers including Lease, Masulis, and Page (1991) point out that the issue date in the SDC database is not very accurate. The problematic cases are offerings launched after the close of stock market trading. For these cases, a researcher should use the data of the next business day. I use the method of Safieddine and Wilhelm (1996) to pick out the actual SEO event date. Their method uses the trading volume surge on the effective issue date around the SEO to identify it. If the next day of an SDC issue date has more than twice the volume of the SDC issue date, and if its volume is more than twice the average daily volume of the previous 250 days, I mark the next day as actual issue date. Corwin (2003) uses the same method, and Altinkilic and Hansen (2003) documents that this method has a high accuracy level.

Table 2 shows some basic statistics about the SEO sample.

Table 2 – Summary statistics of order imbalances

I report summary statistics for SEO characteristics in the dataset. My SEO data contains 1,096 SEOs during 1996 ~ 2004 period. Issuer stock volume is monthly average trading volume measured 6 months from the issue date. I also show SEO sample periods of a few other SEO studies.

Year	Obs.	Statistic	Shares Filed (Million shares)	Amount Filed (Million $)	Issuer Market Value (Million $)	Issuer Stock Trading Volume (Million shares)
1996	11	Mean	2.5	70.7	538.3	12.1
		Median	1.9	34.0	225.1	11.4
1997	138	Mean	4.0	112.5	428.7	8.1
		Median	2.7	62.4	162.7	4.3
1998	58	Mean	8.4	229.8	1428.7	20.5
		Median	3.1	73.4	212.1	8.2
1999	99	Mean	6.7	287.8	2476.9	33.9
		Median	3.0	100.0	202.5	10.5
2000	73	Mean	10.3	281.7	2181.9	28.0
		Median	3.8	113.5	474.6	12.9
2001	165	Mean	20.9	550.16	2507.5	59.0
		Median	4.3	106.6	503.0	15.6
2002	111	Mean	9.6	247.2	2360.0	46.0
		Median	5.0	106.6	809.7	21.7
2003	309	Mean	6.9	158.6	2193.3	61.4
		Median	5.0	89.4	430.0	18.4
2004	132	Mean	5.8	144.3	1454.0	35.6
		Median	4.0	93.9	540.3	18.8
Total	1096	Mean	9.0	241.8	1914.0	42.3
		Median	4.0	88.3	402.1	13.1

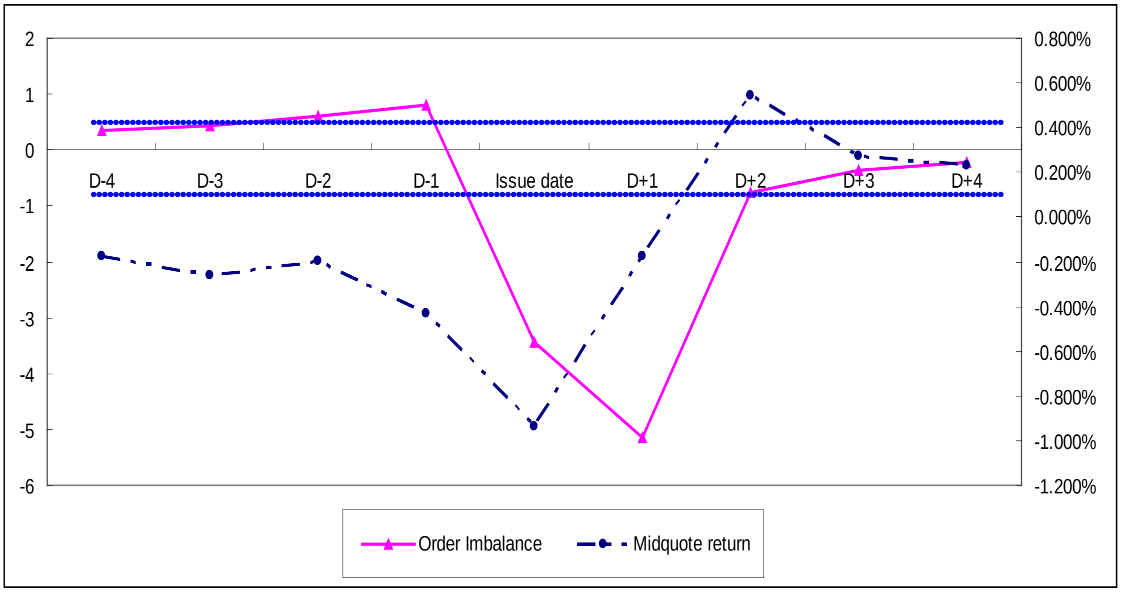
SEO Paper	Sample period
Altinklic and Hansen (2003)	1990 ~ 1997
Altinklic and Hansen (2006)	1985 ~ 2001
Corwin (2003)	1980 ~ 1998
Cotter, Chen, and Kao (2004)	1997 ~ 1998
Meidan (2005)	1993 ~ 2002
Mola and Loughran (2004)	1986 ~ 1999
Safieddine and Wilhelm (1996)	1980 ~ 1991

IV. Order Imbalance around an SEO Issue Date

A. Order imbalance and stock return

Table 3 shows the summary statistics of daily order imbalances and stock returns around SEO issue dates. Order imbalance is normalized by the number of shares outstanding before an SEO and daily stock returns are calculated as a percentage of closing bid-ask mid-points. Figure 2 plots daily order imbalances and stock returns around SEO issue dates.

For this sample, stocks have average daily order imbalance of 0.080 and the standard deviation of the average is 0.378. Order imbalances in [-1, 2] event window around an issue date are particularly significant. The price pressure pattern shows that there is buying pressure before an issue date and selling pressure afterwards. The positive order imbalance before an issue date is consistent with market maker inventory management hypothesis. Market makers lower their quotes to reduce their inventory level before an SEO. However, we can see that order imbalance does not move solely by inventory management because the size of selling pressure after an issue date is much larger than that of previous buying pressure. Observing large selling pressure after SEO issue dates is consistent with price support hypothesis and flipping hypothesis. The former states that underwriters are placing a large limit orders to buy near the offer price, and latter states that institutional investors are selling their allocated stocks. We will be able to differentiate two explanations later using regressions.



- Average order imbalance higher or lower than dotted band is significant in 10% level.

Figure 2 – Order imbalance and stock return around an SEO issue date

Table 3 – Summary statistics of order imbalances

I report summary statistics for order imbalance and stock return around an SEO issue date. Order imbalance is normalized by the number of shares outstanding before an SEO, and stock return is calculated using bid-ask mid-point prices. Order imbalance measures are multiplied by 1000 for visual convenience. Plus or minus indicates business days from an SEO issue date. In my data, average daily order imbalance is 0.080 and standard deviation of the average is 0.378. The mean order imbalances significant in 1% level are marked with small a, the coefficients significant in 5% level are marked with small b, and the coefficients significant in 10% level are marked with small c.

Panel A: Order Imbalance

	Day - 4	Day – 3	Day - 2	Day - 1	Issue D	Day + 1	Day + 2	Day + 3	Day + 4
Mean	0.35	0.45	0.60	0.80[c]	-3.44[a]	-5.15[a]	-0.75[b]	-0.35	-0.21
Median	0.18	0.23	0.30	0.39	0.32	-1.59	-0.09	-0.02	0.08
STD	3.14	3.00	3.31	3.56	20.69	23.07	6.14	4.84	7.17

Panel B: Mid-quote Stock Return

	Day - 4	Day - 3	Day - 2	Day - 1	Issue D	Day + 1	Day + 2	Day + 3	Day + 4
Mean	-0.17%	-0.26%	-0.20%	-0.43%	-0.94%	-0.18%	0.55%	0.28%	0.23%
Median	-0.14%	-0.17%	-0.13%	-0.19%	-0.64%	-0.09%	0.28%	0.17%	0.16%
STD	4.28%	4.13%	4.04%	4.69%	4.47%	4.20%	2.81%	2.49%	2.70%

Panel C: Correlation between Order imbalance and Mid-quote Stock Return

	Day - 4	Day - 3	Day - 2	Day - 1	Issue D	Day + 1	Day + 2	Day + 3	Day + 4
Coefficient	2.33	2.50	2.51	1.39	0.40	0.45	1.16	1.33	1.70
t-value	5.67	5.95	6.92	3.46	5.99	7.67	8.12	8.49	9.32

The pattern of order imbalances is inconsistent with the predictions of both the selling incentives and the uncertainty / information asymmetry hypotheses, while the stock returns evidence is consistent with predictions of both hypotheses. This evidence suggests that shares prices can move without trading pressure (i.e. price jumps to a new level without much trading), or that the effects of two hypotheses are dominated by market maker trading, underwriter price support, and flipping hypotheses.

While the correlations between order imbalances and stock returns should always be positive and significant, there are days with positive order imbalances and negative stock returns. We can also see that while order imbalances generally push stock returns in the same direction, stock returns can move without any large trading pressures. One implication of this result is that stock returns do not always reflect overall trading patterns.

B. Determinants of order imbalance

Since order imbalance can be affected by multiple factors, I use standard regression techniques to identify the determinants of order imbalances and then select variables to test the predictions of each of the five main hypothesis.

Selling incentives related variables:

- Offer size: Corwin (2003) argues that the correlation between offer size and underpricing is evidence of price pressure. Since he measures underpricing by the difference between the offer price and the closing price on the trading day before the issue date, his logic implies that more investors are willing to sell stocks because of the forthcoming supply shock. So order imbalance before an issue date should be more negative as offering size (the size of supply shock) increases. Also, a larger offer size indicates greater opportunities for arbitrage, which starts by short selling the SEO stock prior to the issuance date. Offer size is defined as shares offered divided by the number of shares outstanding before an SEO.[22]

- Kyle's lambda: Kyle's lambda is a sensitivity of a stock to buying or selling pressure. The sensitivity is estimated by regressing stock returns on share order imbalances. The estimation period is 1 year, and I use the sensitivity of 6 months before the issue date. Like offering size, order imbalance would be decreasing in Kyle's lambda.

Uncertainty and information asymmetry related variables:

- Idiosyncratic risk: Moeller, Schlingemann, and Stulz (2006) suggest that information asymmetry can be measured by a stock's idiosyncratic risk, while uncertainty can be represented by stock return volatility. I estimate idiosyncratic risk by regressing daily stock returns on the value weighted market returns taken from the CRSP database. The estimation period is the 6 months prior to the issue date. Cross-sectionally idiosyncratic risk should be negatively correlated with order imbalances prior to the SEO issue date, because information asymmetry will keep buyers from buying SEO stocks. After the issue date, order imbalances may increase in idiosyncratic risk because buyers return to market once the information asymmetry is resolved.

[22] Average stock volume can be an alternative normalizing variable. I find no difference in test results by switching the normalizing variable.

- Stock volatility: Uncertainty is measured by standard deviation of stock return. The estimation period that used in estimating idiosyncratic risk. Cross-sectionally, volatility should have the same correlation with order imbalances as idiosyncratic risk.
- Underwriter reputation: Almost every IPO and SEO study finds that underwriter reputation reduces underpricing (Eckbo, Masulis, and Norli 2007). This result is used as evidence that better underwriters reduce information asymmetry and uncertainty of an equity offering. I follow Altinkilic and Hansen (2003) in measuring underwriter reputation. I measure the market share of the seasoned equity offerings of each investment bank and rank it into 10 categories from worst to best (1 ~ 10). The underwriter reputation is the rank of its market share in SEO and IPO underwritings in the prior year. If an underwriter had no underwritings in the prior year, it receives a 0 value. One problem with the ranking variable is that it is positively skewed. Following Bernard and Thomas (1990) and Mendenhall (2004), the raw rank is divided by 10 and then 0.5 is subtracted to shift the midpoint of the variable to zero.
- Information risk: Bid-ask spread is often used as a proxy for information asymmetry. Glosten and Milgrom (1985) show that spread is increasing in the probability of informed trading. However, Stoll (2000) document that quoted bid-ask spread contains components of information asymmetry, market maker inventory risk, and real frictions. He shows that information asymmetry can be measured by the difference between quoted spread and traded spread. Traded spread is acquired from daily volume-weighted average of bid and ask price. I

classify the difference between the quoted spread and traded spread as asymmetric information risk. Like other variables, information risk is measured over the 6 months before the issue date.

Underwriter price support related:

- Offer size: If offer size is large, the number of shares traded in secondary market increases a lot after an issue date, and as a result, it becomes harder to push stock price in a specific direction. Thus, to support the stock's price, underwriters must place a larger and more frequent limit orders. So order imbalance at or after the issue date will be more negative as the offer size becomes larger. This correlation has a sign that is opposite the prediction of the price pressure hypothesis.
- Underwriter reputation: Cotter, Chen, and Kao (2004) and Lewellen (2006) find that underwriters with a better reputation provide stronger price supports. Thus, order imbalances are expected to be more negative as an underwriter's reputation rises. Although underwriter reputation is also related to information asymmetry and uncertainty, we can differentiate two effects because information asymmetry and uncertainty would affect order imbalance before the issue date, while price support would affect order imbalance at / after the issue date. Also, the predicted correlation is opposite that of information asymmetry and uncertainty hypothesis. Lewellen (2006) reports that price supports after the issue date are uncorrelated with both the information asymmetry and uncertainty measures. Underwriter reputation should also be negatively correlated with order imbalances, because more reputable underwriters are able and willing to undertake stronger price

support activity, which leads to negative order imbalances.

- Prevailing price level: As noted by Cotter, Chen, and Kao (2004), there is little need for price supports if market prices are above the offer prices. Prevailing price level can be estimated using daily average bid and ask prices weighted by trade volume. Stoll (2000) suggests that the mid-point of two prices is a fair estimate of daily prevailing price level. If one day's prevailing price level is lower than offer price, the variable takes a value 1 and is 0 otherwise. This dummy variable is used to explain order imbalances at or after the issue date.

Flipping related:

- Underwriter reputation: Flipping hypothesis predicts that underwriter reputation is positively correlated with order imbalance. Better underwriters would reduce flipping activities by signaling a good performance of an SEO stock. Better underwriters can pressure other institutions to stay away from flipping, using their market power.
- Prevailing price level: To the opposite of underwriter price support hypothesis, prevailing price level dummy will have a positive correlation with order imbalance. The variable receives value 1 when one day's prevailing price level is lower than offer price. Flipping would almost vanish in this case, making order imbalance more positive.

Market maker inventory management related:

- Quote compared to price: Stoll (1978) shows that if market makers want to reduce their inventory levels, they can promote buying pressure by placing quotes below prevailing price levels. To measure this effect, I compare bid and ask quotes with actual trading price. Each day, I subtract average price level from average bid-ask mid point, and this variable is a measure of inventory management. As the bid-ask midpoint falls below the price level, order imbalances become more positive. Thus, the correlation between the price level minus midpoint differences and order imbalances should be negative. This variable is used to explain order imbalances before the issue date.

- Traded bid-ask spread: Inventory risk indicates the risk of market makers incurred by deviating from their optimal inventory level. If the risk is high, market makers would have a larger incentive to smooth their inventory movement around an SEO. Stoll (2000) shows that traded bid-ask spread – the difference between daily average bid and ask – reflects inventory risk and real frictions. Since this spread does not include an information asymmetry component, it is a cleaner measure of inventory risk. I use monthly average of daily traded spreads measured over the prior 6 months from the issue date. Higher traded spreads should make order imbalances more positive prior to the issue date.

A firm's market value is included in the regressions as a standard control variable. The market value is a monthly average measured over the 6 months prior to the issue date. I add a dummy variable that takes a value of 1 when a stock is listed on Nasdaq and is 0 otherwise. Masulis and Shivakumar (2002), Altinkilic and Hansen (2003), and Mola and

Loughran (2004) find that being listed on Nasdaq changes a stock's behavior during an SEO. Table 4 summarizes the variables used in the regressions and their predicted effects on order imbalances.

Table 4 – Explanatory variables

This table shows the summary of explanatory variables and their predicted correlated with order imbalance before/ at/ after an SEO issue date.

Hypothesis	Variable	Before an issue date	At an issue date	After an issue date
Selling incentives	Offer size	Negative	Positive	Positive
	Kyle's Lambda	Negative	Positive	Positive
Uncertainty and information asymmetry	Idiosyncratic risk	Negative		
	Return volatility	Negative		
	Underwriter reputation	Positive		
	Information risk	Negative		
Underwriter price support	Offer size		Negative	Negative
	Underwriter reputation		Negative	Negative
	Prevailing price dummy		Negative	Negative
Flipping	Underwriter reputation		Positive	Positive
	Prevailing price dummy		Positive	Positive
Inventory management	Quote compared to price	Negative		
	Traded spread	Positive		

C. Regression results

The basic approach I take is to estimate a linear equation with order imbalances on the left hand side.

$$OI_{i,t} = \alpha + \mathrm{B} \cdot \mathrm{X}_i + \varepsilon_i \qquad (1)$$

where OI_{it} is order imbalance of stock i, at t days from its issue date and X is matrix of explanatory and control variables

I use OLS with heterskedasticity robust error structure. The error structure is further corrected for clustering by firm. I also add year dummies to control year fixed effect. Table 5 shows regression results.

Table 5 shows that order imbalances before the issue date are mostly uncorrelated with the explanatory variables. The regression results do not detect evidence consistent with the selling incentives or information asymmetry/ uncertainty hypotheses. Moreover, stock volatility is positively correlated with order imbalance 2 days before the issue date, which is directly opposite to the prediction of information asymmetry / uncertainty hypothesis. One implication is that typical negative stock return before an SEO issue date is not a result of SEO related selling activities. Consistent with inventory management hypothesis, quotes compared to prices are negatively correlated with order imbalances in the 3 days and 1 day prior to the issue date. The result supports the conclusion that a positive order imbalance before an issue date is generated by market maker inventory management. However, this type of inventory management has little effect on stock returns or underpricing, because market maker inventory management only moves quotes, rather than prices.

Table 5 –Determinants of order imbalance

Model: $$OI_{i,t} = \alpha + \mathrm{B} \cdot \mathrm{X}_i + \varepsilon_i \qquad (1)$$

I use OLS with heteroskedasticity corrected errors. I account for clustering by firm in the error structure and add year dummies in the regression. P-values are in the parenthesis. The coefficients significant in 1% level are marked with small a, the coefficients significant in 5% level are marked with small b, and the coefficients significant in 10% level are marked with small c.

Panel A: Order imbalance before an issue date

Dependent Variable	Order imbalance 4 days before	Order imbalance 3 days before	Order imbalance 2 days before	Order imbalance 1 day before
Offer size	0.521 (0.25)	0.196 (0.68)	0.141 (0.83)	-0.040 (0.91)
Kyle's Lambda	-0.121 (0.35)	-0.047 (0.74)	-0.167 (0.24)	-0.335[b] (0.02)
Idiosyncratic risk	0.399 (0.74)	-1.685 (0.21)	1.633 (0.30)	2.496 (0.22)
Stock volatility	2.210 (0.74)	22.461[a] (0.00)	0.591 (0.93)	3.143 (0.69)
Information risk	-20.778 (0.61)	-48.737 (0.29)	6.494 (0.92)	-15.835 (0.77)
Underwriter reputation	0.254 (0.50)	0.526 (0.19)	0.427 (0.34)	-0.117 (0.78)
Quote compared to price	2.615 (0.77)	-0.028[b] (0.04)	-3.692 (0.36)	-7.803[b] (0.05)
Traded spread	0.286 (0.82)	-0.536 (0.78)	1.401 (0.34)	-0.602 (0.77)
Log (Firm's Market value)	0.108 (0.30)	0.063 (0.51)	0.015 (0.88)	0.302[a] (0.01)
Nasdaq dummy	-0.795[a] (0.00)	-0.820[a] (0.00)	-1.256[a] (0.00)	-0.642[b] (0.03)
Observations	1078	1076	1079	1079
Adj. R-square	4.9%	5.7%	5.1%	8.1%

Panel B: Order imbalance at or after an issue date					
Dependent Variable	Order imbalance at an issue date	Order imbalance 1 day after	Order imbalance 2 days after	Order imbalance 3 days after	Order imbalance 4 days after
Offer size	-8.450[c] (0.07)	-14.301[a] (0.01)	-1.932[a] (0.01)	0.750 (0.51)	-4.211 (0.14)
Kyle's Lambda	-0.270 (0.81)	-0.212 (0.84)	0.039 (0.92)	-0.234 (0.32)	-0.017 (0.95)
Idiosyncratic risk	3.721 (0.60)	-0.075 (0.99)	3.768 (0.20)	-1.496 (0.28)	0.601 (0.81)
Stock volatility	-101.593[c] (0.07)	6.124 (0.91)	-7.761 (0.55)	11.421 (0.28)	10.693 (0.60)
Information risk	-630.549 (0.18)	422.668 (0.44)	-0.430 (0.99)	-132.298 (0.12)	-82.260 (0.56)
Underwriter reputation	2.966 (0.41)	-8.121[a] (0.00)	-1.722[c] (0.05)	-1.174[c] (0.08)	0.062 (0.96)
Prevailing price dummy	-5.802[a] (0.01)	-6.999[a] (0.00)	-1.431[a] (0.00)	-0.965[a] (0.00)	-0.654[c] (0.09)
Traded spread	0.515 (0.97)	-41.285[c] (0.06)	-2.170 (0.66)	-2.588 (0.53)	0.754 (0.85)
Log (Firm's Market value)	0.497 (0.47)	2.115[b] (0.01)	0.677[a] (0.00)	0.381[b] (0.01)	-0.036 (0.84)
Nasdaq dummy	-0.492 (0.79)	-2.670 (0.24)	-1.252[b] (0.02)	-0.556 (0.19)	-1.285[b] (0.04)
Observations	1080	1080	1080	1080	1080
Adj. R-square	13.4%	10.8%	8.3%	5.7%	4.5%

Note that Nasdaq stocks consistently have a more negative order imbalance before an issue date. Correlation between Nasdaq listing and SEO underpricing is driven by larger selling pressures in the pre-SEO period. There is no clear explanation for why Nasdaq stocks have a greater frequency of negative order imbalances. The explanation for this issue result is a potentially interesting topic for future research.

At the issue date, stock volatility is significantly and negatively correlated with order imbalance levels. The sign of coefficients indicate that investors just react at the equity issue date, rather than altering their trading patterns in anticipation of the event. Meanwhile, order imbalance at an issue date is negatively correlated with offer size and the prevailing price dummy for market price dropping below offer price. Such negative correlations can last for 2 to 3 days after the issue date. Further, underwriter reputation generates significant negative order imbalances after the issue date. This result is consistent with the price support hypothesis that underwriters are placing limit orders to buy stock to prevent the stock prices from dropping much relative to the offer prices. The degree of price support (negative order imbalance) increases in offer size and underwriter reputation. On the other hand, negative coefficient of underwriter reputation and prevailing price dummy is inconsistent with flipping hypothesis. Prevailing price has especially large effect on trading patterns: I find average order imbalance of -8.71 when average market price is below the offer price, while -2.29 for the other cases.

Overall, I find trading patterns around an SEO issue date are mainly driven by market maker inventory management prior to SEO issue dates and underwriter price support following the issue dates. Inventory management has some effects on order imbalances before the issue date, while price support activity has a large impact on order imbalances both at and after the issue date. Other characteristics of an SEO tend to have at most marginal correlations with the share trading patterns.

V. Order Imbalance and SEO Underpricing

This section studies relation between order imbalance and SEO underpricing. The purpose of the test is to see whether trading pressure itself can change the size of underpricing. The selling incentives and information asymmetry / uncertainty hypotheses predict order imbalances are negatively correlated with underpricing in the cross section. This prediction follows from the assumption that selling pressure and information asymmetry / uncertainty generate a negative order imbalance that pushes down the SEO offer price. On the other hand, the underwriter price support hypothesis predicts a positive correlation between order imbalances and underpricing across issuing firms. More negative order imbalance indicates that underwriters are undertaking stronger price support activities, and this activity will help prop up the stock price of an issuing firm.

While SEO underpricing can be defined in several ways, a typical measure is the difference between offering price and the closing price on the SEO issue date. Altinkilic and Hansen (2006) express underpricing in the following equation:

$$underpricing = \log(\frac{P_o}{P_c}) \qquad (2)$$

where P_o is offering price and P_c is the closing price of an issue date.

Since underpricing measures price movement at the issue date, I test the correlation between order imbalances at the issue date and SEO underpricing. Corwin (2003) and Eckbo, Masulis, and Norli (2007) document that variables like offer price, idiosyncratic risk, stock volatility, underwriter reputation, firm size, and Nasdaq dummy are significantly correlated with underpricing. My previous regression on order imbalance contains those variables, and so I use these same control variables plus order imbalance as regressors.

Because order imbalance is correlated with several characteristics of the SEO at an issue date, a regression between underpricing and order imbalance is subject to an endogeneity problem. I use a standard 2-stage-least-squares technique to overcome this problem. In the 1st stage, I estimate order imbalance at the issue date based on SEO characteristics. The 2nd stage regression is between the estimated order imbalances and underpricing. An important part of this estimation is selecting instrument variables in the 1st stage. The instruments should have a high correlation with order imbalances and low correlations with SEO characteristics.

According to Chordia, Roll, and Subrahmanyam (2000) and Chordia and Subrahmanyam (2004), order imbalance has a strong serial correlation. I test serial correlation between order imbalance a day before an issue date and order imbalance at an issue date. I find two variables are significantly and positively correlated (t-stat 3.11). The correlation is significant in 1% level even after controlling for heteroskedasticity in the error structure, clustering by firm, and year fixed effects. Meanwhile, as we see in Table 4, order imbalances before the issue date do not have significant correlations with SEO characteristics, except for Kyle's lambda. Order imbalances a day before the issue date are likely to have a low correlation with SEO underpricing. Also, I show that order imbalances before the issue date is driven by market makers' inventory management, which would not have much effect on stock prices. I verify that order imbalances on the day before the SEO issue dates are not significantly correlated with underpricing (t-stat 1.17).

In the 1st stage equation, I estimate order imbalances on the issue date using the previous day's order imbalance as an instrumental variable. The right hand side has all the explanatory and control variables used in the previous order imbalance regression. The equation is:

$$OI_{i,t} = \alpha + \beta \cdot OI_{i,t-1} + \mathrm{B} \cdot \mathrm{X}_i + \varepsilon_i \qquad (3)$$

I verify that the 1st stage regression model has reasonable explanatory power with an adjusted R-square of 13.1%. 2nd stage regression uses the estimated order imbalances from the 1st stage equation as additional regressors.

$$underpricing_{i,t} = \alpha + \delta \cdot \overline{OI_{i,t}} + \mathrm{B} \cdot \mathrm{X}_i + \varepsilon_i \qquad (4)$$

Wooldridge (2002) shows that this method gives a consistent estimation if X, the matrix of regressors is identical in both regressions. So when altering the model specification, I use the same explanatory variable matrix X in both stages. Table 6 contains the regression estimates for equation (4).

Table 6 –Order imbalance and underpricing

$$\text{Model:} \quad OI_{i,t} = \alpha + \beta \cdot OI_{i,t-1} + \mathrm{B} \cdot \mathrm{X}_i + \varepsilon_i \qquad (3)$$

$$underpricing_{i,t} = \alpha + \delta \cdot \overline{OI_{i,t}} + \mathrm{B} \cdot \mathrm{X}_i + \varepsilon_i \qquad (4)$$

Equation (4) uses OLS with heteroskedasticity corrected errors. I account for clustering by firm in the error structure and add year dummies in the regression. P-values are in the parenthesis. The coefficients significant in 1% level are marked with small a, the coefficients significant in 5% level are marked with small b, and the coefficients significant in 10% level are marked with small c.

Variables	Model 1	Model 2	Model 3	Model 4
Order imbalance	1.192[b] (0.03)	1.356[b] (0.03)	1.462[b] (0.01)	1.315[b] (0.02)
Offer size	0.004 (0.62)	0.006 (0.44)		
Kyle's Lambda	2.881 (0.29)	3.297 (0.25)		
Idiosyncratic risk	2.063 (0.94)	-2.623 (0.93)		
Stock volatility	0.342[a] (0.00)	0.383[a] (0.00)		0.385[a] (0.00)
Information risk	-0.609 (0.67)	-0.517 (0.73)		
Underwriter reputation	-0.022[a] (0.00)	-0.023[a] (0.00)		-0.023[a] (0.00)
Prevailing price dummy	-0.035[a] (0.00)			
Traded spread	-0.052 (0.22)	-0.058 (0.17)		
Log (Firm's Market value)	-2.112 (0.22)	-2.814[c] (0.12)	-6.006[b] (0.02)	-3.219 (0.21)
Nasdaq dummy	7.589 (0.22)	8.405 (0.17)	11.238[b] (0.02)	5.204 (0.28)
Observations	1080	1080	1096	1087
Adj. R-square	13.3%	5.1%	3.5%	4.6%

Table 6 shows that order imbalance is positively correlated with SEO underpricing. The sign of the coefficient is consistent with the price support hypothesis. A stronger price support – more negative order imbalance – reduces underpricing. The correlation between order imbalance and underpricing is significant after controlling for standard determinants of underpricing. Thus, trading pressure alone can affect the degree of underpricing. While underpricing is a common measure of flotation cost, these results show that underwriters can temporarily move secondary market prices to measured flotation costs in terms of underpricing. Large investment banks or commercial banks in particular would have advantage due to their greater ability to move secondary market prices. An implication is that the underpricing measure can be manipulated if a financial institution has ability to move the stock price on the issue date.[23]

VI. Conclusion

Using order imbalance measure, I analyze buying and selling activities around an SEO issue date. I set five hypotheses on order imbalance around an issue date. First, selling incentives hypothesis assumes that traders mostly engage in selling activities before an issue date. There is arbitrage opportunity of short selling before an issue date and covering it afterwards. Supply shock from primary market can create a temporary price drop, and traders may also want to use the opportunity. Second, uncertainty and information asymmetry hypothesis states that uncertainty and information asymmetry of an SEO will keep buyers from trading before an issue date. A result is more sell side trading before an issue date. Third, underwriter price support hypothesis predicts negative

[23] Similarly, Lewellen (2006) finds underwriters with brokerage service engage more in price support activity of an IPO. His result supports the argument that ability to move secondary market is important.

order imbalance after an issue date, because underwriters are placing limit orders at bid side to support stock price. Fourth, flipping hypothesis tells that institutional investors acquire SEO shares in primary market and sell them in secondary market. The activity can give institutional investors a short term profit and generate a selling pressure after an issue date. Fifth, market maker inventory management hypothesis argues that market makers reduce their inventory level before an issue date by lowering their quotes compares to price level. Due to this trading, order imbalances before an issue date become positive.

I track order imbalance patterns around an SEO issue date and find that order imbalance is positive before an SEO and large negative afterwards. The pattern of order imbalance is consistent with price support hypothesis, flipping hypothesis, and inventory management hypothesis. Meanwhile, selling incentives hypothesis and uncertainty / information asymmetry hypothesis have little explanatory power for trading patterns. A factor analysis on order imbalance gives evidence consistent with inventory management hypothesis and price support hypothesis. A positive order imbalance before an issue date is generated by placing quotes lower than price level. After an SEO issue date, selling pressure is increasing in offer size, underwriter reputation, and when current market price drops lower than offer price.

Using an endogeneity controlled regression, I find that price support activity can affect the degree of SEO underpricing. Stocks that received more price support have less underpricing, even after controlling for other known factors of underpricing.

This paper makes two contributes to the literature. First, this paper identifies the main factors that determine stock trading pattern around an SEO. Trading activity is mostly affected by artificial supply and demand in secondary market, such as price support or inventory management. I find no evidence that short selling activity or supply shock from primary market changes overall trading pattern of secondary market. My result shows that trading activity around an SEO issue date is not quite as the same as stock return movement. Second, this paper shows that the degree of underpricing can be manipulated by price support activity. While underpricing is a common measure of equity flotation cost, I find a short term trading activity can affect the size of underpricing. This result raises a question whether underpricing measure well represents equity flotation cost.

References

Altinkilic, Oya and Hansen, Robert S., 2003, Discounting and underpricing in seasoned equity offers, *Journal of Financial Economics 69*, 285-323.

Altinkilic, Oya and Hansen, Robert S., 2006, The puzzling wealth loss around equity offerings: Is it merely price pressure or is there more information?, Working paper, University of Pittsburgh.

Aggarwal, Reena, 2000, Stabilization activities by underwriters after initial public offerings, *Journal of Finance 55*, 1075-1103.

Autore, Don M., 2007, The pricing of seasoned equity offers, Rule 10b-21, and the shelf exemption, Working Paper, Florida State University.

Boehmer, Ekkehart and Fishe, Raymond P. H., 2003, Who receives IPO allocations? An analysis of 'regular investors'. Working paper, University of Miami.

Bernard, Victor L. and Thomas, Jacob K., 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics 13*, 305-340.

Butler, Alexander W., Grullon, Gustavo, and Weston, James P., 2005, Stock market liquidity and cost of issuing equity, *Journal of Financial and Quantitative Analysis 40*, 331-348.

Chemmanur, Thomas J., He, Shan, and Hu, Gang, 2007, The role of institutional investors in seasoned equity offerings, Boston College working paper.

Chordia, Tarun, Roll, Richard, and Subrahmanyam, Avanidhar, 2002, Order imbalance, liquidity and market returns, *Journal of Financial Economics 65*, 111-130.

Chordia, Tarun and Subrahmanyam, Avanidhar, 2004, Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Economics 72*, 485-518.

Corwin, Shane A., 2003, The determinants of underpricing for seasoned equity offers, *Journal of Finance 58*, 2249-2279.

Corwin, Shane A., Harris, Jeffrey H., Lipson, Marc L., 2004, The development of secondary market liquidity for NYSE-Listed IPOs, *Journal of Finance 59*, 2339-2373.

Cotter, James F., Chen, Anlin, and Kao, Lanfeng, 2004, Underwriter price stabilization of seasoned equity offerings: The evidence from transactions data, Academia Economic Papers 32, 53-81.

Eckbo, B. Espen, Masulis, Ronald W., Norli, Oyvind, 2007, Security offerings, *Handbook of Corporate Finance*, edited by B. Espen Eckbo., North-Holland, Chapter 6.

Ellis, Katrina, Michaely, Roni, and O'Hara, Maureen, 2004, Competition in investment banking: Proactive, reactive, or retaliatory?, Working paper, University of California.

Glostern, Lawrence R. and Milgrom, Paul R., 1985, Bid, ask and transaction in a specialist market with heterogeneously informed traders, Journal of Financial Economics 14, 71-100.

Habib, Michael A., Ljungqvist, Alexander P., 2001, Underpricing and entrepreneurial wealth losses in IPOs: Theory and evidence, Review of Financial Studies 14, 433-458.

Henry, Tyler R. and Koski, Jennifer L., 2008, Short selling around seasoned equity offerings, Working paper, University of Georgia.

Huh, Shan-Wook and Subrahmanyam, Avanidhar, 2005, Order flow patterns around seasoned equity offerings and implications for stock price movements, *International Review of Finance 5*, 75-111.

Kadlec, Gregory B., Loderer, Claudio F., Sheehan, Dennis P., 1994, Issue day effects for common stock offerings: Causes and consequences, Working paper, Penn State University.

Kim, Kenneth A. and Shun, Hyun-Han, 2004, The puzzling increase in the underpricing of seasoned equity offerings, *Financial Review 39,* 343-365.

Kraus, Alan and Stoll, Hans R., 1972, Price impacts of block trading on the New York Stock Exchange, *Journal of Finance 27*, 569-588.

Kriegman, Laurie, Shaw, Wayne H., Womack, Kent L., 2001, Why do firms switch underwriters?, *Journal of Financial Economics 60*, 245-284.

Lease, Ronald C., Masulis, Ronald W., and Page, John R., 1991, An investigation of market microstructure impacts on event study returns, *Journal of Finance 46*, 1523-1536.

Lee, Charles M.C., and Ready, Mark J., 1991, Inferring trade direction from intraday data, *Journal of Finance 46*, 733-747.

Lewellen, Katharina, 2006, Risk, reputation, and IPO price support, *Journal of Finance 61,* 613-663.

Masulis, Ronald W., Shivakumar, Lakshmanan, 2002, Does market structure affect the immediacy of stock price response to news?, *Journal of financial and quantitative analysis 37*, 617-648.

Mendenhall, Richard R., 2004, Arbitrage risk and post-earnings-announcement drift, *Journal of Business 77*, 875-894.

Meidan, Danny, 2005, A re-examination of price pressure around seasoned equity offerings, Working Paper, Northwestern University.

Mikkelson, Wayne H., and Partch, M. Megan, 1985, Stock price effects and costs of secondary distributions, Journal of Financial Economics 14, 165-194.

Moeller, Sara B., Schlingemann, Frederik P., 2006, Does investor diversity of opinion, information asymmetry, or uncertainty resolution affect acquirer returns?, Working paper, Ohio State University.

Mola, Simona and Loughran, Tim, 2004, Discounting and clustering in seasoned equity offering prices, *Journal of Financial and Quantitative Analysis 39*, 1-23.

Myers, Stewart C., and Majluf, Nicholas S., 1984, Corporate financing and investment decisions when firms have information that investors do not have, *Journal of Financial Economics 13*, 187-221

Petersen, Mitchell A., 2007, Estimating standard errors in finance panel data sets: comparing approaches, *forthcoming Review of Financial Studies.*

Safieddine, Assem and Wilhelm, William J., 1996, An empirical investigation of short-selling activity prior to seasoned equity offerings, *Journal of Finance 51*, 729-749.

Scholes, Myron, 1972, The market for corporate securities: Substitution versus price pressure and the effects of information on share price, *Journal of Business 45*, 179-211.

Stoll, Hans R., 2000, Friction, *Journal of Finance 55*, 1479-1514.

Stoll, Hans R., 1978, The supply of dealer services in securities markets, *Journal of Finance 33*, 1133-1151.

Wooldridge, Jeffrey M., 2002, *Econometric analysis of cross section and panel data*, MIT press.

Appendix – Construction of order imbalance data

1. Criteria for stock selection are:

- Data source comes from Trade and Quote (TAQ) data.
- Data period is from January 1996 to December 2004.
- I exclude Certificates, ADRs, shares of beneficial interest, units, Americus Trust components, closed-end funds, preferred stocks and REITs from the dataset.
- I delete the stock is from the sample year if the price at any month-end during the year was greater than $999.
- I eliminate non-synchronous trading issue by marking stock return as missing if there was no trade on today or previous day.

2. When constructing order imbalance variable, I only use quotes and trades such that:

- Quotes and trades are in regular market trading times (from 9:30 to 16:00)
- There is no special settlement conditions
- All bid-ask spreads are positive

3. Method to calculate order imbalance is (Lee and Ready (1991) method):

- A trade is buyer (seller) initiated if it is closer to the ask (bid) of the prevailing quote.
- Prevailing quote should be at least 5 seconds old.

If the trade is at the midpoint of the quote, the trade is buyer (seller) initiated if prior stock price change was positive (negative).