An Investigation of Information Content in Stock Order Flows

By

Tao Zhang

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

(Business)

at the

UNIVERSITY OF WISCONSIN – MADISON

2013

Date of final oral examination: 08/27/2013

The dissertation is approved by the following members of the Final Oral Committee:
Kenneth Kavajecz, Professor, Business
Mark Ready, Professor, Business
Elizabeth Odders-White, Associate Professor, Business
Terry Warfield, Professor, Business
Bruce Jones, Professor, Agriculture Economics

Acknowledgement

I am very grateful to my adviser, Professor Kenneth Kavajecz, for his great mentorship, constant encouragement and exceptional guidance, and to the rest of my dissertation committee—Professor Mark Ready, Professor Elizabeth Odders-White, Professor Terry Warfield and Professor Bruce Jones—for their valuable inputs and insightful advices. Also I would like to express my gratitude to Belle Heberling and Danielle Nguyen, and Phil Miller for their help throughout the program and the warmth of their supportive personality.

I would like to thank my parents for their faith in me and allowing me to pursue whatever dreams I have had in my life. They have taught me ample wisdom of life. Finally, I am thankful to my girlfriend Sissi for her support, encouragement, and unwavering love. She is a big part of the reason why I think all my hard work is worthwhile.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Daily stock trading has taken place at a higher frequency since the introduction of electronic exchanges by the U.S. Securities and Exchange Commission (SEC) in 1998. High frequency trading allows economic information to be incorporated into stock prices at a faster speed through competition among high frequency traders. How and at what speed the economic information is incorporated into stock prices are two interesting questions for market microstructure researchers. The availability of high frequency transaction data on stock quotes and trades makes it possible for researchers to examine how information is reflected in investors' trading and stock prices. In contrast, low frequency data such as daily stock prices and returns usually show where the market ends up but does not contain information about what paths stock prices take. For example, the end-of-day price may be driven by a small trade at the end of the day while stock order flows should have more accurate descriptions about the overall stock performance. As a result, stock order flows should reflect investors' trading activities on a more accurate basis. Transaction data is supposed to have richer information content than low frequency data and in addition should be particularly useful if researchers are interested in the connection between investors' trading activities and stock market movements. Prior to the availability of transaction data, trading volume data was analyzed by researcher in order to study the connections between returns and trading activities. The information content in trading volume, however, can be very different from that in stock order flows. Trading volume data lacks information on trade direction (buyer-initiated or seller initiated). Volume is widely considered as a proxy for stock liquidity and thus may have very different implications for stock prices than transaction data.

The intra-day trading data usually contains information on each trade: trading direction (with orders signed as the well-known Lee and Ready, 1991 algorithm), trading volume and transaction price. Buyer-initiated orders, usually signed as positive, are trades with transaction prices higher than the mid-point of the prevailing bid and ask spreads. Transaction data reveals more detailed information about investors' trading pattern and therefore attracts significant research attention. Most of the attention has been focused on exploring these data for extra information that is not contained in low frequency data and using such information to predict the performance of stock market and macro economy. A great deal of research has studied the connections between stock returns and net order flows (e.g. Chordia and Subrahmanyam, 2004; Chordia Roll, and Subrahmanyam 2005). These studies find that order flows are positively related to contemporaneous and future returns. The positive relation as pointed out by these studies may be attributed to the positive autocorrelations in order flows. The positive autocorrelations in order flows suggest that on average investors tend to trade in the same direction due to the fact that information spreads gradually and investors react at different speeds to economic news.

The objective of my dissertation is to seek answers to the more general question of how information is processed and incorporated by the market. In particular, my study emphasizes the role of stock liquidity in the process of information spread. The importance of liquidity can be justified by simple economic intuition. Liquid stocks would be a less costly choice for investors to trade in response to economic news and therefore should have quicker reaction than illiquid stocks. Hence if we examine movements of order flows of liquid stocks we should be able to extract useful information about the future performance of illiquid stocks and the whole market. The role of liquidity in information dissemination, although very important, was rarely discussed in prior research. My study contributes to the literature in two ways. First, I show that liquidity has significant influence on stock price reaction to economic news. Second, I show that order flows of liquid stocks contain useful information to predict values of illiquid stocks and market states. My dissertation consists of two different yet internally connected studies. In particular, the first study investigates how investors trade after earnings announcements. My research provides empirical evidence that stock order flows provide additional information about stock values after public earnings announcements. I choose earnings announcements as the information event for three reasons. First, the post-earnings announcement drift (PEAD) is a famous asset pricing anomaly and has attracted significant research attention. Understanding how information spreads and investors trade after earnings announcements will help us discover the correct value of mispriced stocks and provide meaningful contribution to the PEAD literature. Second, the earnings announcements have been shown to have information spillover effects on stock returns in the same industry. A firm's earnings reports usually convey useful information about its peer firms and the prospect of the whole industry. Investors have the option to trade multiple stocks after earnings announcements. This spillover effect makes earnings announcement a good candidate for us to study investors' trading activities in stocks with different liquidities exposed to economic news. Third, there are complete datasets available on each earnings announcement. These datasets provide information on announcement dates and times, consensus analysts' forecasts and actual earnings, which allows me to examine performance of stock order flows and the returns around different earnings surprises. The results of my research are consistent with the hypothesis that investors prefer to trade more liquid stocks after earnings announcements and the order flows in liquid stocks contain very useful information for the prices of announcing firms' stocks.

The second study focuses on comovements of stock order flows, which are typically incurred by the arrival of new information on the market. Commonality of order flows is a good measure of comovements of investors' trading reaction to information shocks. Thus it plays an important role in helping researchers understand how information is incorporated into stock trades and prices as well as what affects the speed of this process. The objective of this study is to investigate whether stock liquidity is an important factor affecting the degree of stock comovements. Specifically, I examine if stocks with higher liquidity experience higher commonality in order flows in response to economic news and if the comovements in order flows of liquid stocks can be used as leading economic indicators that provide us with useful information about the market prospect. Intuitively, sophisticated investors would prefer to allocate their funds to more liquid stocks when they expect market conditions to deteriorate and would more likely to invest in illiquid stocks when they expect market conditions to turn up. The results of this study show that liquidity has a very strong positive effect on commonality of stock order flows. Stocks with higher liquidity generally have higher order flows commonality and are more responsive to economic news than low liquidity stocks. Moreover, an increase in commonality among large (small) stocks predicts an increase (decrease) in market uncertainty one or two months ahead. This implies that informed investors move to large and liquid stocks when the market liquidity is going to decline and the information uncertainty is going to rise.

My research presents empirical examples of how we can acquire important information about asset prices and the stock market by analyzing high frequency data. The richness of transaction data has been enhanced by some recent research efforts classifying trades into transactions by large traders and small traders. Institutional investors are usually more sophisticated investors with trading strategies supported by statistical models while retail investors' trades are more based on fundamental news and investors' sentiment. Therefore their trading behaviors convey quite distinct information and have different implications on asset prices. For example, Hvidkjaer (2008) classified trades into small investors' trades and large investors' trades by the trade size (dollar volume). The screening methodology based on statistical rules, although not perfect, provides researchers with important methods to analyze institutional and retail investors' behaviors separately.

My findings convey the importance of relative liquidity in information dissemination and price discovery. Sometimes it is the relative liquidity of a stock compared to its peer stocks that affect investors' trades when investors' have special preference for a group of stocks that have some common characteristics. For example, institutional investors may prefer value/growth stocks, or large/small stocks during some periods in time because their models tell them to do so. Investors that try to time the sectors may be particularly interested in stocks of a certain sector (consider the IT bubble period). Therefore the relative liquidity of a stock should play an important role in the cross section of asset returns. There has been little research conducted on relative liquidity and its effect on trading and asset prices. A close examination of relative liquidity will contribute to current asset pricing literature both theoretically and empirically.

Chapter 2

The Post-Earnings-Announcement Drift and Order Flows of Correlated Stocks

Abstract

The post-earnings-announcement drift (PEAD) states that the immediate stock price reaction to public earnings announcements does not fully reflect the true value of the stock, leaving the rest of the value to be discovered by investors' trading activities. Past studies have found that the earnings announcement has a significant and positive spillover effect on the returns of stocks in the same industry, which implies that investors with a better understanding of stock values may choose to invest in these stocks if there is less trading friction. I investigate the informational role of stock order flows of nonannouncing firms in the same industry and document three empirical findings: (1) the stocks of nonannouncing firms experience drifts similar to the stocks of announcing firms after an earnings announcement, especially after a positive earnings surprise announcement. (2) The stock order flows of nonannouncing firms in the same industry relative to the order flows of announcing firms have significant explanatory power for the PEAD of the stocks of announcing firms after controlling for a rich set of conventional explanatory variables. (3) Illiquidity, as a measure of trading friction, explains a significant part of investors' trading preference during the announcement window. Moreover, consistent with my hypothesis, the relation between net order flows and the PEAD is most pronounced in the subsample where event stock has the lowest liquidity compared to its industry peers.

1. Introduction

The post-earnings-announcement drift (PEAD) first documented by Ball and Brown (1968) has presented a long-standing challenge to the market efficiency hypothesis. The PEAD refers to the fact that, in general, companies that report good earnings news (positive earnings surprise) tend to have long-run positive abnormal stock returns after the announcement and companies that report bad earnings news (negative earnings surprise) tend to have negative abnormal returns. Figure 1 is excerpted from Bernard and Thomas (1989, Fig. 2) and presents a simple illustration of this anomaly. The stock price experiences an initial jump after the news is first published. Then, in the post-announcement period, the stock returns exhibit an abnormal drift in the same direction as the initial response. In a more general sense, recent studies refer to this anomaly in terms of the fact that good news announcers tend to outperform bad news announcers in the post-announcement period. Chordia and Shivakumar (2006) find that a portfolio long in stocks with the highest earnings surprises and short in stocks with the lowest earnings surprises generates a monthly return of 0.9 percent or more than 10 percent annually. In a semi-strong form efficient market, any strategies that are based on past or public information should not be profitable.

Based on two stylized facts that trading friction plays an important role in explaining the PEAD and that the earnings announcement has significant and positive spillover effect on the returns of stocks in the same industry, this paper investigates the informational role of stock order flows of nonannouncing firms in the same industry. I ask the question: Do trades in correlated stocks of nonannouncing firms contain information about the value of the stocks of announcing firms?

I first demonstrate that there is a drift of the same sign in the stock returns of nonannouncing firms after the earnings announcement, especially for positive earnings surprises.

The drift in stock returns of nonannouncing firms provides evidence that sophisticated investors have incentives to invest in these stocks. I then collect the order flows data from the Institute for the Study of Security Markets (ISSM) and the Trades and Quotes (TAQ) dataset, measure investors' trading interest in stocks of nonannouncing firms relative to the event stock by the difference between abnormal net order flows of these stocks, and test the relation between this measure and the PEAD directly. Empirical results suggest that, stock order flows of nonannouncing firms in the same industry have significant and positive explanatory or predictive power for the PEAD of the stocks of announcing firms. High net order flows (more buyer-initiated order flows) in correlated stocks of nonannouncing firms relative to the stocks of announcing firms are associated with higher drift in stock returns of announcing firms. This relation is robust and significant after controlling for a rich set of conventional explanatory variables. To test whether investors' trading preference is influenced by trading friction, I examine how a stock's liquidity compared to its industry peers drives abnormal order flows over the short event window and find that stocks with higher liquidity typically have a larger order imbalance. The relation between order flows and the PEAD is most pronounced in the subsample where event stocks have the lowest liquidity compared to their industry peers. Finally, my study results indicate that order flows contain more information than stock returns in explaining the PEAD.

My study makes several important contributions to the existing literature. First, a similar drift discovered in stock returns of nonannouncing firms in the same industry accounts for over 50 percent of the drift in stock returns of announcing firms, suggesting that part of the PEAD is likely to be associated with risk compensation to rational investors for that industry, which is a new explanation pertinent to the PEAD. My study also finds that sophisticated investors may have incentives to invest in stocks of nonannouncing firms upon receiving earnings news.

Illiquidity as a measure of trading friction affects investors' preference over these stocks. Although previous studies point out that illiquidity and other transaction costs impede price adjustment of stocks of announcing firms, I explicitly demonstrate that market friction actually drives investors' trading preference in stocks exposed to the earnings news, which supports the role of trading friction in explaining the PEAD documented in past studies and enhances our understanding of this market anomaly. Moreover, to my knowledge, my paper is the first that describes the role played by information externality and order flows in predicting and explaining the PEAD and opens an interesting area for PEAD research. More importantly, it brings up an interesting and broader issue: Since order flows contain important information regarding price discovery and market efficiency, how should we utilize the information in order flows more effectively? More research is needed to better understand how order flows of different stocks work together. Finally, the relation between order flows and PEAD revealed in this work may help investors make better investment decisions and thus contribute to market efficiency.

The rest of the paper is organized as follows. Section 2 reviews related literature. Section 3 investigates the role of correlated stocks of non-announcing firms and develops theoretical hypotheses. Section 4 discusses the data used for empirical analysis and the sample selection procedure. Section 5 demonstrates the existence of PEAD in my sample and reveals a similar drift in the stock returns of nonannouncing firms. Section 6 reports the results of univariate and multivariate tests that were performed to examine the relation between relative net order flows and the PEAD. Section 7 investigates how market friction captured by illiquidity is related to investors' trading activities and the relation between order flows and PEAD documented in Section 6. The final section provides a discussion and conclusion.

2. Literature Review

The persistence and robustness of the PEAD anomaly have attracted much attention among researchers. Francis et al (2007) find that stocks of announcing firms with greater private information have more muted initial market reactions and larger subsequent returns. Chordia and Shivakumar (2005) suggest that the PEAD is related to macroeconomic conditions and that the drift may be explained by macroeconomic factors such as inflation. Mendenhall (2004) uses the component of a stock's risk that cannot be hedged as a measure of arbitrage risk and shows that trading strategies based on earnings surprises are subject to high arbitrage costs. Sadka (2006) decomposes liquidity risk into variable and fixed components and finds that the variable component is priced in the context of the PEAD portfolio returns. He argues that liquidity risk can explain a significant portion of the cross-sectional variation of the PEAD portfolio returns. Some studies focus on investors' behavior patterns. Recent studies find that the PEAD is related to investors' limited attention. The main idea of this line of research is that people's minds are finite, and investors cannot fully react to earnings news if they are distracted. For instance, DellaVigna and Pollet (2009) hypothesize that investors are more likely to be distracted on Fridays (they call this the "weekend distraction effect") and find that Friday announcements have a 15 percent lower immediate response and a 70 percent higher delayed response than weekday announcements. Hirshleifer, Lim, and Teoh (2009) argue that multiple simultaneous earnings announcements will distract investors and cause them to underreact to earnings news.

Another strand of literature focuses on the trading friction and transaction cost of forming the PEAD portfolio. Since the abnormal return from the PEAD portfolio is short-lived, the implementation of this strategy involves high portfolio turnover, which leads to high transaction costs. Bhushan (1994) documents a stronger drift for smaller, low-priced stocks whose transaction costs are supposed to be higher than the transaction costs of larger stocks. Hou and Moskowitz (2005) demonstrate that the PEAD is prevalent in stocks that have the most friction as measured by the delay by which their prices adjust to information. Ng, Rusticus, and Verdi (2008) examine the effect of transaction costs on the PEAD directly and find weaker return responses at the time of the earnings announcement and higher subsequent returns drift for firms with higher transaction costs. Chordia et al. (2009) examine the profitability of the PEAD portfolio for stocks with different liquidity levels. They document that the PEAD occurs mainly in highly illiquid stocks whose transaction costs are relatively high. They conclude that transaction costs explain 70 percent to 100 percent of the paper profits from a long–short strategy designed to exploit the earnings momentum anomaly. Therefore, trading friction keeps investors who have superior valuation of stocks from trading away the drifts.

It is worth noting that the main focus of my study is not to provide another alternative explanation for the long-lasting PEAD anomaly although I point out that the drift in stock returns of non-announcing firms may serve as evidence of risk compensation for rational investors dedicated to that industry. In fact, my study is related to the explanation based on trading friction and tries to uncover extra information contained in stock order flow to make better prediction the PEAD.

3. Earnings Announcement and Correlated Stocks

Earnings announcements have externality effects. Once an earnings announcement is made public, it is not uncommon for there to be an information transfer from announcing firms to non-announcing firms. For example, the following report of IBM's earnings announcement on January 20, 2009, notes an unexpected earnings increase among other firms.

> *IBM shares up 4.1% to $85.30 after the company said its fiscal fourth-quarter profit was $4.4 billion, or $3.28 share, on revenue of $27 billion…Analysts surveyed by FactSet Research were looking for earnings of $3.03 a share on revenue of $27.9 billion for the quarter ended in December…Separately, late-traded shares of Hewlett-Packard Co. rose*

2.3% to $34.10 and Oracle Corp. gained 2.2% to $16.45. Cisco Systems Inc. was also higher, up 1.3% to $15.20. (Wall Street Journal, January 20, 2009)

As shown in this example, an unexpected earnings increase at one firm leads to an overall appreciation in the stock prices of the firm's competitors. The information externality effect of earnings announcements has been widely studied as well. For example, Foster (1981) documents a strong impact of a firm's earnings announcement on the stock prices of other firms in its industry. Han and Wild (1990) find a significant positive relation between unexpected earnings of announcing firms and unsystematic stock returns of nonannouncing firms. Wasley (2001) finds that cross-sectional variation in nonannouncing firms' stock returns over the announcement period is explained by the unexpected component of the preannounced earnings of a firm in the same industry. Thomas and Zhang (2008) confirm that a firm's earnings announcement has a strong positive spillover effect on the stock returns of other firms in the same industry and show that investors of these nonannouncing firms tend to overreact to announcing firms' earnings announcements. Although one can argue that a positive earnings surprise in the stocks of announcing firms may imply either good or bad news for stocks of nonannouncing firms, all of these studies indicate that, in general, the effect is statistically positive.

Based on the information externality of earnings announcement and the fact that trading friction impedes rapid price discovery, this paper investigates where sophisticated investors prefer to trade after an earnings announcement in the presence of trading friction and how their trading activities provide information about the PEAD. Sophisticated investors are defined as those who have superior analytical skills of stock values after learning the earnings announcement. Because earnings announcements usually convey information about the performance of firms in the same industry, if sophisticated traders choose to invest in the stocks of nonannouncing firms as well, we should expect order flows in these stocks to have information content for the future stock price (the PEAD) of the announcing firm. The idea of my study is similar to that of Easley, O'Hara, and Srinivas (1998), who investigate where investors prefer to trade if they have private information about a stock. They find that option volume plays an important informational role in predicting future price movements of the corresponding stock. In particular, positive (negative) option trades such as buying (selling) a call or selling (buying) a put predict upward (downward) price movement in the stock market. They argue that factors involving liquidity, option leverage, and the fraction of informed investors may make one market more attractive to investors than another. Similarly, I examine whether market friction affects sophisticated investors' preference for stocks. If stocks of nonannouncing firms are more attractive to these investors, then trading activities in these stocks may first reflect information common to the corresponding industry and have predictive power for the price movement of the event stock. Figure 2 illustrates this idea.

To support this argument, I test the following two hypotheses:

H1: Order flows of stocks of nonannouncing firms provide positive explanatory power toward the PEAD.

H2: Investors prefer to trade in more liquid stocks in the same industry during the announcement window.

To test these hypotheses, I examine the high-frequency order flows data. Order flows data are considered more closely related to trades based on private valuation and stock price discovery. The stylized fact that investors underreact to earnings announcements implies that on the announcement day, the market adjusts quickly to the public information released but fails to reflect all the information perfectly, leaving the rest of price discovery to trading based on private valuation, which is associated with order flows imbalance.

4. Data and Sample Selection

Quarterly earnings announcement data were collected from the CRSP-Compustat merged database and I/B/E/S from 1983 to 2005. When these two datasets contain a discrepancy regarding the earnings announcement date, I follow DellaVigna and Pollet (2009), who do random checks on the accuracy of the announcement dates from these two datasets and report that the earlier of the two dates is almost always the correct announcement date. To exclude stale forecasts, the most recent one-quarter-ahead forecasts are required to be less than 60 days old. To measure the earnings surprise, I calculate the deviation of the true earnings from the most recent record of the analysts' consensus forecasts, which is defined as the median of the forecasts from individual analysts. Then the earnings surprise of firm i in quarter q is defined as the difference between the true earnings and the prevailing forecast scaled by the stock price at the end of the corresponding quarter:

$$ES_{iq} = \frac{e_{iq} - F_{iq}}{P_{iq}},$$

where i and q are firm and quarter indices, respectively, and e is the true earnings reported by the firm. F is the most recent consensus forecast, and P is the stock price at the end of quarter q. Earnings or forecasts with absolute value greater than the stock price or stock whose price is less than $1 are excluded from the sample to minimize data errors. To evaluate firms' performance, I collect daily stock return data from CRSP and calculate the cumulative abnormal returns (CAR) of postannouncement windows. Following DellaVigna and Pollet (2009) and Hirshleifer, Lim, and Teoh (2009), I allow for a two-day event window [0,1]—the announcing day and the following day—for investors to make an initial reaction to the earnings announcement. If the announcement is made after the trading hours, then the next day is considered as the event day. Then the postannouncement CAR is defined as the difference between the buy-and-hold stock return of the announcing firm and that of a size and book-to-market (B/M) matching portfolio over the long-term window, starting from the next day after the announcement window [0,1] and covering the subsequent 60 trading days [2,61] relative to the announcement window. I also compute the CAR over the event window and use it as a control variable for the regressions. Returns are split-adjusted. Specifically, the cumulative abnormal returns over the event window and the postannouncement window are computed as:

$$CAR\ [0,1]_{iq} = \prod_{j=t}^{t+1}\left(1 + R_{ij}\right) - \prod_{j=t}^{t+1}\left(1 + R_{pj}\right)$$

$$CAR\ [2,61]_{iq} = \prod_{j=t+2}^{t+61}\left(1 + R_{ij}\right) - \prod_{j=t+2}^{t+61}\left(1 + R_{pj}\right)$$

where R_{ij} is the return of firm I, and R_{pj} is the return of the size and B/M matching portfolio on day j, where t is the announcement date of quarter q's earnings.

The selection of 60 days is based on Bernard and Thomas's (1989) findings that most of the drift occurs during the first 60 trading days after the announcement. For each stock, I use 1 of the 25 size and B/M portfolios to which it belongs as its matching portfolio. The daily returns of the 25 portfolios are available from Kenneth French's data library website.[1] The market capitalization of each stock is obtained at the end of June. The book-to-market ratio is computed as the book equity of the last fiscal year-end in the prior calendar year divided by the market value of equity at the end of December of the prior year.

The intraday order-level data are collected from the Consolidated Quotes and Consolidated Trades datasets from the ISSM (1983–1992) and the TAQ (1993–2005) dataset. Since the information externality effect is an important component of my study, I focus on medium-sized industries that have 5 to 15 firms identified by their four-digit SIC code. This

[1] See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

procedure is similar to procedures applied in previous studies. For example, Foster's (1981) sample consists of industries with 6 to 12 firms, and Wasley (2001) requires the selected industries to have at least 5 and no more than 15 competitors. This screening process allows me to have enough nonannouncing observations to average out noise order flows without information content as well as earnings announcements influential to the corresponding industry. For each announcing firm, I find four nonannouncing firms that have the closest market capitalization to the announcing firm in the same four-digit SIC industry. There are 5 stocks in each event group. These stocks are required to have daily return data in the CRSP dataset and intraday quotes and trades data in that month of the announcement. Since I include a set of control variables, the observations in my sample require data availability on these control variables as well. My final sample consists of 51,099 firm-announcement observations covering 3,844 firms. About 32.77% (16,746 observations) of the sample has negative earnings surprise, 53.21% (27,190 observations) of the sample has positive earnings surprise, and 14.02% (7,163 observations) has zero earnings surprise.

5. The Drift of Announcing and Nonannouncing Firms

This section discusses the post-earnings-announcement performance of announcing and nonannouncing firms in terms of stock returns. I first demonstrate that the PEAD occurs in my sample and is more apparent for stocks with positive surprise. Then I show that the stocks of nonannouncing firms have abnormal returns similar to those of announcers. This finding suggests that sophisticated investors may prefer to invest in the stocks of nonannouncing firms if these stocks have less trading friction.

5.1 The Post-Earnings-Announcement Drift

Ball and Brown (1968) point out that the stock market tends to underreact to earnings announcements and that there is an abnormal drift in stock returns in the same direction as the initial reaction to the news. Much subsequent research has confirmed the persistence and robustness of this finding using updated data and measures of abnormal returns. In this section, I demonstrate that the PEAD occurs in my sample. I split the sample into three groups: positive surprise, negative surprise, and no surprise. I then compute the average CAR over the [2,61] window for each group. The CARs [2,61] calculated for each of these groups are presented in Panel A of Table 1. The PEAD is positive and significant for the positive-surprise group (1.56% with a t statistic of 20.44). For the negative-surprise group, there is a significant negative drift (−0.61% with a t statistic of −6.33). These post-announcement drifts confirm the findings of prior research regarding the PEAD. In my sample, the no-surprise group also has a small negative drift (−0.40% with a t statistic of −2.5).

5.2 The Drift in Stock Returns of Nonannouncing Firms

If sophisticated investors have better valuation of stocks and choose to invest in correlated stocks of nonannouncing firms as well, it can be expected that stocks of nonannouncing firms will have similar drifts after an announcement to attract these investors. I compute the cumulative abnormal returns for an equal-weighted portfolio of stocks of nonannouncing firms over periods of 5, 15, 30, 45, and 60 days after the announcement window (i.e., the windows [2,6], [2,16], [2,31], [2,46], and [2,61]) following either good earnings news or bad earnings news. I compare these drifts to those in stock returns of announcers for the good-news group and bad-news group, respectively. The average drifts calculated for both announcing and nonannouncing firms are shown in Panel B of Table 1. During the announcement window [0,1], both the stocks of announcing and nonannouncing firms experience abnormal returns of the same sign as the earnings surprise. This is consistent with previous findings that earnings announcements have positive spillover effects on stock returns in the same industry. Not surprisingly, the abnormal returns are significantly larger in stocks of announcing firms (1.79%

versus 0.1% for the positive-surprise group and (−1.97% versus −0.03% for the negative-surprise group). The equal-weighted portfolio of stocks of nonannouncing firms has drifts of the same sign as the stocks of announcing firms for the positive surprise group. The magnitude of the drifts is larger for the stocks of announcing firms in the long run. It is interesting that after controlling for size, book-to-market ratio, and industry, the size of the drifts in stock returns of nonannouncing firms accounts for over 50 percent of the drifts in stocks of announcing firms, which implies that some of these drifts may be attributed to risk compensation to rational investors for similar-sized firms in that industry.

However, I find that the long-term drifts over these windows after negative surprises are not significant. This asymmetry suggests that unexpected earnings increases contain more information about the firms in the same industry, while unexpected earnings decreases are more likely to be interpreted as firm specific rather than as applying to an entire industry. In fact, past studies point out some systematic differences in information content of positive surprises and negative surprises. Basu (1997) documents that unexpected earnings increases have higher persistence than unexpected earnings declines in that the conservatism principle results in greater timeliness of bad news. In other words, an unexpected earnings decrease is not as informative of the firm’s own earnings as unexpected earnings increases, and therefore it is not surprising that an unexpected earnings decrease contains less information for other firms’ earnings in the same industry. In a more recent study, Ranasinghe (2012) finds that the stock price reacts positively to nonnegative earnings surprises that confirm earlier earnings news in the same industry, but there is no such effect for confirmatory earnings with negative surprises. This finding is consistent with mine: negative earnings surprises are not as informative as positive ones signaling the industry prospect.

Figure 3 plots the average cumulative abnormal returns over the announcement window and postannouncement period for both announcing and nonannouncing firms. It can be observed from the figure that the stocks of announcing firms experience a stronger announcement window CAR than the stocks of nonannouncing firms. However, after the announcement window, the cumulative abnormal returns for the announcing and nonannouncing firms are closer in size for the positive surprise group. For the negative surprise group, there does not appear to be a significant negative drift in stock returns of nonannouncing firms. The drift for nonannouncing firms after positive earnings surprises suggests that sophisticated investors have incentive to invest in these stocks after learning good news about the industry.

The analysis in this section provides evidence for my claim that stocks of nonannouncing firms in the same industry may be an optional venue for sophisticated investors due to a similar drift in these stock returns. On the announcement day, the investors who realize that current stock prices have not adjusted fully to the earnings news may choose to invest in either stocks of announcing firms or correlated nonannouncing firms.

6. The PEAD and Order Flows

This section examines the informational role of stock order flows of nonannouncing firms in predicting the post-announcement performance of announcing firms' stocks. I calculate the normalized net stock order flows of both announcing and nonannouncing firms and use the difference between the two to capture the extra information content in the stock order flows of nonannouncing firms. Univariate and multivariate tests are performed to investigate the relation between the relative net order flows measure (RNOF) and the PEAD.

6.1 The Relative Net Order Flow Measure

Following most of the research on the information externality of earnings announcements, I focus on investors' trading activities and order flows behavior of stocks of interest during the two-day event window [0,1][2] and examine whether their order flows convey extra information about the PEAD of the stocks of announcing firms. The quote and trade data from 1983 to 1992 are available from the ISSM dataset, and data from 1993 to 2005 are available from the TAQ dataset. These two datasets provide intraday quotes and trades data of all stocks listed on the NYSE, AMEX, and NASDAQ back to 1983. Since there is no record about whether a trade is buyer-initiated or seller-initiated, I match these quotes and trades and infer trade direction following the standard five-second rule (Lee and Ready, 1991).[3] If multiple bids and asks occur at the same second, I use the last record as the prevailing price at that time. The net order flows of a given period are defined as the difference between total buyer-initiated dollar (share) volume traded and the total seller-initiated dollar (share) volume traded within that period. To capture the unexpected or abnormal order imbalance, the total net order flows occurring in a day is demeaned by the corresponding daily average net order flows of that month and scaled by the standard deviation of net order flows in that month; that is:

$$SNOF_{it} = \frac{NOF_{it} - MOF_{im}}{SD_{im}},$$

where $SNOF_{it}$ is a standardized net order flows measure for stock i on day t, NOF_{it} is the net order flows of stock i on day t, MOF_{im} is the mean of the net order flows of stock i in announcement month m, and SD_{im} is the standard deviation of net order flows of stock i in month m.

Finally, I compute the average of the standardized daily net order flows during the two-day announcement window and obtain the net order flows measure for a stock. To measure the

[2] For announcements made after trading hours I compute order flows over the window [1,2].

[3] I also use three-second, one-second, and no-time-adjustment rules to infer the trade direction as a robustness check. My results remain the same.

amount of extra information captured in stock order flow of nonannouncing firms, for each event stock I define a measure of relative net order flows as:

RNOF = average standardized net order flows of nonannouncing firms – standardized net order flows of the announcing firm.

This variable captures investors' trading interest in peer stocks of nonannouncing firms relative to the event stock conditional on the earnings news. The normalization by standard deviation allows us to look at abnormal order flows movement independent of a liquidity effect. Under my hypothesis H2 that order flows of nonannouncing firms provide additional information about the value of the event stock, a positive relation between this variable and the PEAD of the event stock should be observed given a positive information externality effect documented in previous research.

6.2 Univariate Tests

Table 2 reports the summary statistics of the relative net order flows measures based on share volume and dollar volume. These two measures have similar distributions and a high correlation coefficient of 0.998. For the share volume measure, the distribution has a mean of slightly less than zero (−0.028) and a standard deviation of 0.912. The minimum is −2.655, and the maximum is about 2.963. For the dollar volume measure, the distribution is almost the same, with a slightly smaller maximum and minimum.

Recall that hypothesis H2 suggests a positive relation between the relative net order flows and the PEAD. In performing a univariate test of this hypothesis, I first sort observations of announcing firms into deciles by earnings surprise. Then I sort observations into terciles (high, medium, and low) independently by the RNOF measure. For each earnings surprise decile, I compute the mean postannouncement period cumulative abnormal returns for the high-RNOF group and for the low-RNOF group and the difference in these cumulative abnormal returns.

Table 3 reports the CARs for the high- and low-RNOF terciles based on dollar volume after controlling for earnings surprise.[4] For eight out of ten earnings surprise deciles, the high-RNOF group earns higher cumulative abnormal return than the low-RNOF group. Three of ten deciles have significant difference between the high- and low-RNOF groups at the 10% level and one at 5% level. This indicates that the abnormal order flows in peer stocks of the same industry carry important information about the stocks of announcing firms. Consistent with the positive spillover effect of earnings announcement documented in the literature, in general when there is more positive (negative) abnormal stock order flows of nonannouncing firms relative to announcing firms, announcing firms are more likely to have higher (lower) abnormal returns in the following 60 days.

6.3 Multivariate Tests

To control for other possible factors that affect the PEAD of announcing firms' stock returns, I perform a multivariate regression including a set of conventional explanatory variables. Previous research shows that investors' reaction to earnings news is related to firm size, book-to-market ratio, earnings persistence, earnings volatility, the number of analysts following the stock, reporting lag, institutional ownership, and share turnover. I include these control variables and let them interact with the earnings-surprise decile in regressions (3) and (6). Size and book-to-market effects are included as size and book-to-market quintiles with breakpoints defined in Section 4. The reporting lag is the number of days from the end of the corresponding quarter until the actual earnings announcement date. The number of analysts following a stock is defined as the logarithm of (1 + the number of active analysts).

Table 4 presents the estimation result under different model specifications with the relative net order flows measures based on both share volume and dollar volume. Because the

[4] RNOF based on share volume generated similar results

relation between the postannouncement CAR and the earnings surprise is highly nonlinear (e.g., Kothari, 2001), following past research, I use the decile rank of earnings surprise as opposed to the earnings surprise itself. The variable EPD is the earnings surprise decile (an EPD of 1 indicates lowest earnings surprise; an EPD of 10 indicates highest earnings surprise). In regressions (3) and (6), I also include indicator variables for year, month, and whether a day is Friday (DellaVigna and Pollet, 2009, find that investors are more likely to be distracted on Fridays). To account for the distraction effect of competing news reports as documented in Hirshleifer, Lim, and Teoh (2009), in regressions (3) and (6) I sort observations into deciles according to the number of earnings announcements on the announcement day and include an interaction term between the number-of-announcements decile and the earnings-surprise decile as they do. The variable NRANK is the number-of-announcements decile rank. To investigate whether the relative net order flows measure contains extra information that is not included in stock returns, I add to the regression the cumulative abnormal return over the two-day event window [0,1]. Standard errors of regression coefficient estimates are adjusted for heteroskedasticity and clustering by the day of announcement. The results suggest that the RNOF is significant (at the 1% level and 5% level for full specification regression) on the cumulative abnormal returns (see Table 4). All coefficients in Table 4 are expressed as percentages. The relation is significant under all specifications with order flows based on share volume and dollar volume. The estimation results based on the dollar-volume order flows measure and on the share-volume order flows measure are very close, with the same R-squared. The positive coefficient (0.707 in regressions (3) and (6)) is consistent with the result in the preceding section and consistent with the hypothesis H2 that the net order flows of nonannouncing firms reveal information that is not fully reflected in the event stock's trading event after controlling for other PEAD determinants.

The distraction effect ESD*NRANK is significant and positive. This is consistent with the results in Hirshleifer, Lim, and Teoh (2009) that the distraction effect leads to a larger PEAD. It is worth noting that the RNOF coefficient in this study is statistically more significant than the distraction effect coefficient. The short-term cumulative abnormal return CAR[0,1] turns out to be insignificant and does not provide much explanatory power for the PEAD. The inclusion of the control variables does not change dramatically the estimation results in regressions (2) and (5), which implies that the relation found between RNOF and PEAD is robust.

In sum, the results of the multivariate regression analyses are consistent with the result of the univariate test and provide evidence in favor of the positive relation between relative net order flows and the PEAD. This relation is robust and significant after controlling for a rich set of relevant variables, which suggests that net order flows of nonannouncing firms over the two-day event window add new information to the price discovery of event stocks that is not contained in past analysis.

6.4 Early and Late Announcements

Early and late earnings announcements may contain different industry-wide information content. While early announcements are supposed to contain more information and incur greater stock price reaction, researchers (e.g. Ramnath 2002) have found that investors are not able to fully appreciate the information from the early announcers' news and this underreaction leads to predictable stock returns for subsequent announcers in the same industry. However, Thomas and Frank (2008) provide different conclusion and show that investors overestimate the intra-industry implications of early announcers' earnings reports and stock returns for late announcers in response to their own earnings news are negatively correlated with those in response to early announcements from their competitors.

To examine if the informative role of order flows depends on the timing of announcements, I rank stocks within each event group by the announcement time of their earnings reports for the corresponding fiscal quarter. Then I examine the explanatory power of order flows for the subsamples where the firm is the first announcer and the last announcer respectively. Interestingly, I find the predictability of stock returns does not particularly depend on the timing of announcement and the explanatory power of order flows is almost symmetric for early and late announcements in the industry.

7. Liquidity and Trading Activities

Based on the preceding argument, a stock's trading friction is an important feature that influences its price discovery during the event window. In this section, I examine stock liquidity as a measure of trading friction and how it is related to the findings in Section 6. I measure liquidity using the Market Quality Index (MQI) suggested by Bollen and Whaley (1998) and the Amihud (2002) price impact measure. A directly estimated transaction cost is also a popular measure of market friction. For example, Keim and Madhavan (1997) use institutional trading data to estimate both implicit and explicit transaction costs. Implicit costs mostly refer to the price impact of trades, and explicit costs refer to direct costs such as commissions. However, I avoid using estimated transaction cost to measure a stock's trading friction, because the direct transaction cost estimation is usually executed by fitting linear regressions using an estimated coefficient from Keim and Madhavan (1997), which implies that transaction costs estimated for a large sample have satisfactory accuracy, but an estimation for an individual stock is likely to be subject to large errors. Stocks in the same event group are ranked by their liquidity later on in Section 7.1. Therefore, the use of estimated transaction cost for each stock is not reliable and may generate misleading results when I do the ranking procedure. Other commonly used transaction cost measures such as quoted and effective bid-ask spread are also considered as liquidity measures. These measures, however, do not contain as much information regarding market depth as the MQI measure. Thus, I will focus on the MQI measure and the Amihud price impact measure. Section 7.2 revisits the relation between the RNOF and the PEAD for groups with event stocks of different liquidity. In Section 7.3, I include stock returns of nonannouncing firms and examine how they help predict the PEAD.

7.1 Liquidity and Order Imbalance

As previously discussed, stock order flows of nonannouncing firms have information content for price discovery of the stocks of announcing firms. If there is friction in trading the event stock, it is expected that there are more trades happening to other correlated stocks of nonannouncing firms with less friction when the fundamental values of these stocks are signaled by the earnings announcement of the event stock. Therefore, market friction should influence investors' trading preference when they learn about the new information. Liquidity is one important and popular measure of market friction. Roughly speaking, liquidity reflects the degree to which an asset or security can be bought or sold rapidly in the market with minimum loss of value and without affecting the asset's price significantly. Liquidity is a multidimensional concept, and different measures of liquidity refer to different aspects characterizing easiness or difficulty in trading a stock.

To measure the friction of trading a stock, I employ two measures of liquidity. I use the MQI suggested by Bollen and Whaley (1998) to measure the quality of stock liquidity. Beber, Brandt, and Kavajecz (2007) show that this variable is more parsimonious compared to other popular measures. MQI is the ratio of the average share depth at the prevailing bid and ask price quotes to the percentage quoted spread.

$$MQI = \frac{(\text{depth at bid} + \text{depth at ask})/2}{\text{percent spread}},$$

where depth at bid and depth at ask are measured in thousands of shares, and percent spread is the quoted bid-ask spread divided by the bid-ask price midpoint. A higher MQI implies higher market depth or tighter spread, and less difficulty in trading a stock at a given price. The advantage of this measure over other measures is that it focuses on the quote data and incorporates both spread and depth information. I compute the average MQI during the two-day event window for both announcing and nonannouncing firms.[5]

The second liquidity measure is the Amihud (2002) average daily price impact of trading volume, which is computed as the absolute price change per dollar of daily trading volume. I compute the liquidity of all stocks during the two-day event window as the average daily liquidity measure:

$$ILQ_{i,t} = \frac{1}{2}\left(\frac{|R_{i,t}|}{Dvol_{i,t}} + \frac{|R_{i,t+1}|}{Dvol_{i,t+1}}\right),$$

where t is the announcement date, $R_{i,t}$ is the return of stock i on day t, and $Dvol_{i,t}$ is the dollar trading volume for stock i on day t. The Amihud (2002) price impact measure actually captures illiquidity of a stock. A higher ILQ suggests that a given amount of trading will cause a greater price impact and that a stock is less liquid. Hasbrouck (2009) compares effective and price-impact measures estimated from daily data with those estimated from intraday data and points out that the Amihud measure has the highest correlation with trade-based measures. The application of these two measures allows me to look at a stock's liquidity estimate based on both quotes and trades data.

[5] Investors' trading patterns may depend on a stock's historical liquidity. I look at the historical liquidity measure using the average liquidity of past three month and find that it is highly correlated with the announcement window liquidity measure. I obtain similar results.

To measure investors' trading interest in a stock, I first compute the standardized net order flow as described in Section 6, which is the daily net order flows demeaned and then normalized by the standard deviation of the net order flows of the stock in the month of the announcement. The normalization is important since unnormalized net order flows measures are correlated to the liquidity measures (for example, more liquid stocks have more dollar trading volume and thus their daily net order flows has higher standard deviation, and it would not be surprising if we find more liquid stocks have more net order flows in dollar volume) and therefore prevent us from identifying the relation between abnormal order flows movements and stock liquidity in the context of an earnings announcement. I use the standardized order imbalance, that is, the absolute value of the average daily standardized net order flows as a measure of investors' trading interest in a stock over the two-day event window.

I investigate within each event group how investors' trading activity is affected by the stock's comparative liquidity. A simple linear regression of the standardized order imbalance on the liquidity measure may fail to capture the underlying relation between order flows and liquidity, because there is great variation in stock's liquidity across event groups (for example, stocks in some groups are larger in size with higher liquidity). Therefore, a stock's relative liquidity to its group members should be more important for interested investors when they make decisions about where to trade. Due to this concern, I rank the stocks within each event group according to their liquidity. A rank number of 1 (5) represents lowest (highest) liquidity in that group (that is, it has the lowest MQI or highest Amihud illiquidity measure in that group). By the Amihud measure, about 11.58 percent of event stocks have lowest liquidity, and about 23.17 percent have highest liquidity. The mean rank is approximately 3.3, and the median is 3. The ranking based on the MQI measure has a similar distribution. I rank the stocks within each group according to the order imbalance to reflect investors' trading interest in a stock relative to other stocks. Again this procedure allows me to control for other factors that may affect trading activities across groups (for example, for some groups the earnings news is more influential and therefore causes higher order imbalance in all stocks). A rank number of 1 (5) represents lowest (highest) order imbalance, implying lowest (highest) trading interest by investors. Then I compute the pairwise simple correlation coefficients between the liquidity rank (based on both the MQI and Amihud measures) and the order imbalance rank. I report the order imbalance rank based on dollar volume since the standardized order imbalance measures in terms of dollar volume and share volume are highly correlated (with a correlation of 0.998) and generate the same result. The pairwise correlation coefficients and the p value for the test of significance are shown in Panel A of Table 5. All the correlation coefficients are positive and highly significant with a p value of approximately zero. The liquidity ranks based on the MQI and Amihud measures are positively correlated with a correlation of 0.512, suggesting high commonality in these two measures despite the fact that they focus on the quotes side and trades side, respectively. The correlation coefficients between liquidity rank and order imbalance rank are about 0.116 and 0.213, respectively, suggesting that in each event group, higher liquidity stocks are associated with higher abnormal order flows and thus respond more sensitively to the public information.

To estimate the effect of the liquidity rank on order imbalance, I try to be conservative and use the MQI liquidity measure, which has a lower rank correlation with liquidity measure.[6] I run linear regression of the order imbalance on the liquidity rank based on the MQI measure. I add the absolute value of earnings surprise to control for the information influence of the earnings announcement. I also include the event dummy variable to indicate whether the stock belongs to an announcing firm and the interaction term of the event dummy variable and

[6] Using the Amihud measure generates more significant results

the liquidity rank to specifically examine the liquidity effect on the event stock's abnormal order flows. The estimation results are summarized in Panel B of Table 5. The coefficient estimated on the liquidity rank is highly significant from zero with a t statistic of 19.37. In contrast, the coefficient on absolute earnings surprise does not have much explanatory power for order flows after controlling for liquidity rank. The positive and significant coefficient on the event dummy variable suggests that, in general, the event stock receives more order imbalance than the stocks of nonannouncing firms during the event window. Liquidity rank has a significant effect on the event stock, and higher liquidity of the event stock attracts more abnormal order flows. In particular, the estimation result suggests that a unit increase in the liquidity rank of the event stock is associated with a 0.02 standard deviation increase in the standardized order flows imbalance.

7.2 Liquidity Rank and the Relation between PEAD and RNOF

Another testable hypothesis is that the relation between the relative net order flows measure and the PEAD found in Section 6.3 is more pronounced for event stocks with smaller liquidity rank (i.e., less relative liquidity). I divide the sample into groups according to the MQI liquidity rank of the event stock. With the same control variable set, regressions (3) and (6) are rerun for each group. Due to the uneven distribution of liquidity rank (there are far fewer observations of stocks of announcing firms with the lowest and highest liquidity rank), to make the estimation results comparable across groups, I combine the two groups with the lowest liquidity ranks of event stock (liqRank = 1 and liqRank = 2) into one group. Similarly, the two groups with the highest liquidity ranks (liqRank = 4 and liqRank = 5) are combined into one group.

Table 6 reports the estimated coefficients for each group. Consistent with the hypothesis above, the coefficient on the relative net order flow measure RNOF is only significant in the subsample that has the most illiquid event stock (liqRank = 1 and liqRank = 2). This relation holds for both regressions based on order flows measured by dollar volume and share volume. The coefficient on the RNOF suggests that the RNOF has the largest effect (0.293 for the share volume regression and 0.295 for the dollar volume regression) on the PEAD and is significant at the 10 percent level in the group with liquidity rank equal to 1 or 2 and not significant for all other groups with higher liquidity event stocks. This finding provides evidence to support my hypothesis H3 that stock liquidity plays an important role in explaining the trading activity of sophisticated investors and the interaction between RNOF and PEAD. Only when trading in an event stock is relatively costly for active investors upon receiving the earnings news, the information contained in order flows of nonannouncing firms conveys information about the event stock's value.

Since the postannouncement drift for stocks of nonannouncing firms is mostly prevailing in positive surprise groups, if sophisticated investors have better valuation of stocks and their trading activities in stocks of nonannouncing firms provide information about stock value of announcing firms, we would expect the relation between RNOF and PEAD to be more pronounced in the positive surprise group. Table 7 displays the relation in the context of different earnings surprises. The sample is divided into positive-, negative-, and no-earnings-surprise groups. Consistent with the above hypothesis, with the same set of control variables, the relation is most significant for the positive-earnings-surprise group with an estimated coefficient 0.24 (0.26 for dollar volume-based estimation). The coefficients are not significant for the negative- or no-earnings-surprise groups. The results are almost identical for dollar-volume-and share-volume-based order flows measures.

7.3 The PEAD and Stock Returns of Nonannouncing Firms over Event Window

We have seen that stock order flows of nonannouncing firms provide information about the stock value of announcing firms, and then it is natural to ask whether stock returns of nonannouncing firms contain similar information as order flows. Thus, I compute average cumulative abnormal returns for nonannouncing firms over the announcement window in each event group and examine their explanatory power for the PEAD. Table 8 presents the estimation results using CAR[0,1] for nonannouncing firms as the explanatory variable. The variable ACAR[0,1] is the average cumulative abnormal returns for nonannouncing firms in each group over the announcement window. The variable RCAR is defined as the difference between ACAR[0,1] and the event stock CAR[0,1], ACAR[0,1]-CAR[0,1], which is analogous to the relative net order flows measure. Neither of these two return variables shows significant explanatory power for the drift of the stocks of announcing firms. The order flow measure, RNOF, remains positive and significant at the 5 percent level, which implies that order flows, as argued, contain more information than stock returns and can be used to predict the PEAD of the stocks of announcing firms.

In sum, this section describes how I use illiquidity as a measure of trading friction to investigate the relation between trading friction and investors' trading preference in the context of earnings announcements. My analysis shows that the comparative difficulty to trade the event stock leads to more abnormal order flows to the correlated stocks of non-announcing firms and therefore makes these order flows more informative about the value of the event stock. Order flows of correlated stocks contain more information and have predictive power for the PEAD of the event stock when there is more friction in trading the event stock.

8. Concluding Remarks

This paper investigates the long-standing post-earnings-announcement drift anomaly by looking at the informational role of stock order flows of nonannouncing firms in the same industry. Based on two stylized facts that the earnings announcement has significant and positive spillover effect on the returns of stocks in the same industry and that trading friction plays an important role in explaining the PEAD, I ask the question: Do trades in correlated stocks of nonannouncing firms contain information about the value of the stocks of announcing firms?

I demonstrate that there is a similar drift in nonannouncing firms' stock returns, mainly for the positive-earnings-surprise group, which accounts for more than 50 percent of the drift in the stocks of announcing firms returns. This suggests that sophisticated traders may have an incentive to invest in these stocks of nonannouncing firms as well. I also investigate how stock net order flows of nonannouncing firms during the short announcement window help explain the PEAD for announcing firms. After controlling for the earnings surprise and a rich set of conventional explanatory variables, I find that the relative net order flows of nonannouncing firms in the same industry have significant explanatory and predictive power for the PEAD in stock returns of announcing firms. I provide empirical evidence that illiquidity, as a measure of friction, explains a significant part of investors' trading preference during the announcement window. Moreover, consistent with my hypothesis, the relation between RNOF and the PEAD are more pronounced in the subsample where event stocks have the most trading friction (lowest liquidity) in the event group. Also consistent with the drift pattern of nonannouncing firms' stock returns, this relation is more pronounced for the positive-earnings-surprise group. Order flows contain more information than stock returns to explain the PEAD.

My study makes several contributions to research in this area. First, it uncovers a similar drift in stock returns of nonannouncing firms in the same industry. This finding suggests that part of the PEAD is likely to be associated with risk compensation for that industry to rational investors, which is a new explanation related to the PEAD. Second, I provide evidence that investors have incentives to invest in stocks of nonannouncing firms in the same industry, and their trading preference is closely related to stock relative liquidity. Although previous studies emphasize that illiquidity and other transaction costs impede stock price adjustment, I explicitly demonstrate that market friction actually drives investors' trading preference in stocks exposed to the earnings news, which enhances our understanding of this market anomaly. Third, to my knowledge, this is the first study that points out the role played by information externality and order flows in predicting and explaining the PEAD. An interesting and broader question is how we should utilize the information in order flows when we understand that order flows contain important information regarding price discovery and market efficiency. More research is needed to better understand how order flows of different stocks work together. Finally, my study suggests that information contained in the order flows of correlated stocks has predictive power for the PEAD of the stocks of announcing firms. This relation may help investors make better investment decisions and thus contribute to market efficiency.

References

Ali A., L. Hwang, and M.Trombley. 2003. Arbitrage Risk and the Book-to-market Mispricing. *Journal of Financial Economics* 69: 355–373.

Amihud, Y. 2002. Illiquidity and Stock Returns: Cross-Section and Time-Series Effects. *Journal of Financial Markets* 5, no. 1: 31–56.

Asquith, P., P. Pathak, and J. Ritter. 2005. Short Interest, Institutional Ownership, and Stock Returns. *Journal of Financial Economics* 78: 243–276.

Ball, R., and P. Brown. 1968. An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research* 6, no. 2: 159–178.

Basu, S. 1997. The Conservatism Principle and the Asymmetric Timeliness of Earnings. *Journal of Accounting and Economics* 24, no. 1: 3–37.

Beber, Alessandro, Michael W. Brandt, and Kenneth A. Kavajecz. 2007. Flight-to-quality or Flight-to liquidity? Evidence from the Euro-area Bond Market. *Review of Financial Studies* 22, no. 3: 925–957.

Bernard, V. L., and J. K. Thomas. 1989. Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium? *Journal of Accounting Research* 27, no. 1:

Bhushan, R. 1994. An Informational Efficiency Perspective on the Post-Earnings Announcement Drift. *Journal of Accounting and Economics* 18, no. 1: 45–65.

Bollen, N., and R. Whaley. 1998. Are "Teenies" Better? *Journal of Portfolio Management* 25: 10–24.

Brennan, M., T. Chordia, and A. Subrahmanyam. 1998. Alternative Factor Specifications, Security Characteristics, and the Cross-Section of Expected Stock Returns. *Journal of Financial Economics* 49, no. 3: 345–373.

Chordia T., A. Goyal, G. Sadka, R. Sadka, and L. Shivakumar. 2009. Liquidity and the Post-Earnings-Announcement Drift. *Financial Analysts Journal* 65, no. 4: 18–32.

Chordia, T., R. Roll, and A. Subrahmanyam. 2001. Market Liquidity and Trading Activity. *Journal of Finance* 56, no. 2: 501–530.

Chordia, T., and L. Shivakumar, 2005, Inflation illusion and post-earnings-announcement drift, *Journal of Accounting Research,* 43(4), 521-556.

Chordia, T., and L. Shivakumar. 2006. Earnings and Price Momentum. *Journal of Financial Economics* 80, no. 3: 627–656.

Daniel, K., and S. Titman. 1997. Evidence on the Characteristics of Cross-Sectional Variation in Stock Returns. *Journal of Finance* 52, no. 1 (March): 1–33.

DellaVigna, S., and Pollet, J., 2009. Investor Inattention and Friday Earnings Announcements. *Journal of Finance* 64: 709–749.

Easley, D., M. O'Hara, and P. Srinivas. 1998. Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. *Journal of Finance* 53: 431–465.

Fama, E. F., and K. R. French. 1992. The Cross-Section of Expected Stock Returns. *Journal of Finance* 47, no. 2: 427–465.

Fama, E. F., and K. R. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33, no. 1: 3–56.

Foster, G. 1981. Intra-industry Information Transfers Associated with Earnings Releases. *Journal of Accounting and Economics* 3: 201–232.

Francis, J., D. Pagach, and J. Stephan. 1992. The Stock Market Response to Earnings Announcements Released During Trading Versus Nontrading Periods. *Journal of Accounting Research* 30: 165–184.

Francis, J., R. Lafond, P. Olsson, and K. Schipper. 2007. Information Uncertainty and

Post-Earnings-Announcement-Drift. *Journal of Business and Accounting 34: 403-433*

Han, J., and J. Wild. 1990. Unexpected Earnings and Intra-industry Information Transfers: Further Evidence. *Journal of Accounting Research* 28: 211–219.

Hanna, D., and M. Ready. 2005. Profitable Predictability in the Cross-Section of Stock Returns. *Journal of Financial Economics* 78, no. 3: 463–506.

Hasbrouck, J., and D. Seppi. 2001. Common Factors in Prices, Order Flows, and Liquidity. *Journal of Financial Economics* 59: 383–411.

Hasbrouck, J., 2009 Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data. *Journal of Finance* 64, 1445–1477.

Hirshleifer, D., S. Lim, and S. Teoh. 2009. Driven to Distraction: Extraneous Events and Underreaction to Earnings News. *Journal of Finance* 6, no. 5: 2289–2325.

Hou, K., and T. Moskowitz. 2005. Market Frictions, Price Delay, and the Cross-Section of Expected Returns. *Review of Financial Studies* 18, no. 3: 981–1020.

Jegadeesh, N., and S. Titman. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* 48, no. 1: 65–91.

Keim, D., and A. Madhavan. 1997. Transaction Costs and Investment Style: An Interexchange Analysis of Institutional Equity Trades. *Journal of Financial Economics* 46, no. 3: 265–292.

Korajczyk, R., and R. Sadka. 2004. Are Momentum Profits Robust to Trading Costs? *Journal of Finance* 59, no. 3: 1039–1082.

Kothari, S. P. 2001. Capital Market Research in Accounting. *Journal of Accounting and Economics* 31: 105–231.

Lee, C., and M. Ready. 1991. Inferring Trade Direction from Intraday Data. *Journal of Finance* 46: 733–754.

Mendenhall, R. 2004. Arbitrage Risk and the Post-Earnings-Announcement Drift. *Journal of Business* 77, no. 6: 875–894.

Nagel, S. 2005, Short Sales, Institutional Investors and the Cross Section of Stock Returns. *Journal of Financial Economics* 78: 277–309.

Ng, J., T. Rusticus, and R. Verdi. 2008. Implications of Transaction Costs for the Post–Earnings Announcement Drift. *Journal of Accounting Research* 46, no. 3: 661–696.

Pástor, L., and R. F. Stambaugh. 2003. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy* 111, no. 3: 642–685.

Ranasinghe, T. 2012. Intraindustry Information Transfers: An Analysis of Confirmatory and Contradictory Earnings News. Working Paper, University of Houston.

Sadka, R. 2006. Momentum and PEAD Anomalies: The Role of Liquidity Risk. *Journal of Financial Economics* 80, no. 2: 309–349.

Thomas, J., and F. Zhang. 2008. Overreaction to Intra-industry Information Transfers? *Journal of Accounting Research* 46, no. 4: 909–940.

Wasley, C. 2001. Information Externalities Associated with Earnings Pre-announcements. Working Paper. University of Rochester.

Wood, A., T. McInish, and K. Ord. 1985. An Investigation of Transaction Data for NYSE Stocks. *Journal of Finance* 25: 723–739.

Table 1
Average Cumulative Abnormal Returns (CARs) for Announcing and Nonannouncing Firms for Different Earnings Surprise Groups

Panel A: The Average Post-Earnings-Announcement-Drift by Earnings Surprise

	CAR[2,61]	t-stat
Earnings Surprise>0	1.56%***	20.44
Earnings Surprise=0	-0.40%**	-2.50
Earnings Surprise<0	-0.61%***	-6.33

Panel B: Drift in Announcing and Non-announcing Stock Returns

	Announcing					
	CAR[0,1]	CAR[2,6]	CAR[2,16]	CAR[2,31]	CAR[2,46]	CAR[2,61]
Good News	1.79%***	0.18%***	0.56%***	0.79%***	1.05%***	1.56%***
Bad News	-1.97%***	-0.20%***	-0.19%***	-0.38%***	-0.67%***	-0.61%***
	Non-announcing					
Good News	0.10%***	0.14%***	0.31%***	0.45%***	0.61%***	0.81%***
Bad News	-0.03%***	0.11%***	0.14%***	0.02%	-0.01%	0.08%

Note: The CAR is computed as the cumulative abnormal returns over different windows relative to the announcement date. CAR is defined as the difference between the buy-and-hold return of the announcing/nonannouncing firm and that of a size and book-to-market matching portfolio over the long-term window, starting from the next day after the announcement window [0,1] and covering the subsequent 60 trading days [2,61] relative to the announcement window. * Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level.

Table 2
Summary Statistics of Relative Net Order flows Measure and Firm Characteristics

Panel A: Distribution of the Relative Net Order flows (RNOF), Share Volume

Mean	Median	Standard Deviation	Min	Max	Obs
-0.028	-0.043	0.912	-2.655	2.963	71346

Panel B: Distribution of the Relative Net Order flows (RNOF), Dollar Volume

Mean	Median	Standard Deviation	Min	Max	Obs
-0.035	-0.042	0.913	-2.335	2.731	71346

Panel C: Firm Characteristics

	Min	Max	Mean	Median
Size	0.10	13.31	6.10	5.96
Book to Market	0.00	77.99	0.61	0.47
Turnover	0.20	5721949.00	134.32	81.56
# of Analyst	1.00	44.00	5.77	4.00

Note: The RNOF measure is computed using order-level data from 1983 to 2005 from ISSM and the TAQ dataset. RNOF = average standardized daily net order flows of nonannouncing firms – standardized net order flows of the announcing firm over the announcement window. Trade direction is inferred following the standard five-second rule (Lee and Ready, 1991). Size is the logarithm of the market value of equity. Turnover is the average quarterly trading volume in thousands.

Table 3
The PEAD of Announcing Stocks by Earnings Surprise and RNOF

	Average CAR[2,61] for high and low RNOF tercile		
Earnings Surprise Decile	RNOF Low	RNOF High	High-Low
1	-2.96%	-1.54%	1.42%*
2	-0.72%	-0.80%	-0.08%
3	-0.83%	-1.67%	-0.84%
4	-0.77%	0.46%	1.23%**
5	-0.38%	-0.30%	0.08%
6	-0.14%	0.85%	0.99%*
7	0.10%	0.14%	0.04%
8	0.12%	0.85%	0.73%
9	1.83%	2.27%	0.44%
10	1.57%	3.06%	1.49%*
10-Jan	4.53%***	4.6%***	

Note: The postannouncement CAR is defined as the difference between the buy-and-hold return of the announcing firm and that of a size and book-to-market matching portfolio over the long-term window, starting from the next day after the announcement window [0,1] and covering the subsequent 60 trading days [2,61] relative to the announcement window. The RNOF measure is computed using order-level data from 1983 to 2005 from ISSM and the TAQ dataset. RNOF = average standardized daily net order flows of nonannouncing firms – standardized net order flows of the announcing firm over the announcement window. Trade direction is inferred following the standard five-second rule (Lee and Ready, 1991). Decile = 1 includes firms with the lowest earnings surprises meaning most negative, and Decile=10 includes the highest meaning most positive. * Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level.

Table 4

Regression of cumulative abnormal returns on relative net order flows and other control variables

	Share Volume			Dollar Volume		
	(1)	(2)	(3)	(4)	(5)	(6)
	CAR[2,61]	CAR[2,61]	CAR[2,61]	CAR[2,61]	CAR[2,61]	CAR[2,61]
ESD	0.418***	0.203***	0.707***	0.418***	0.203***	0.707***
	(0.037)	(0.085)	(0.220)	(0.037)	(0.085)	(0.220)
RNOF	0.249***	0.246***	0.240**	0.252***	0.250***	0.246**
	(0.081)	(0.085)	(0.100)	(0.081)	(0.085)	(0.100)
CAR[0,1]		-0.061	-0.200		-0.060	-0.200
		(0.185)	(0.187)		(0.185)	(0.187)
NRANK*ESD		0.025***	0.015		0.025***	0.015
		(0.010)	(0.014)		(0.010)	(0.014)
Controls and controls interacted with earnings surprise decile			YES			YES
constant	-2.165	-2.156**	-7.292***	-2.165	-2.155**	-7.291***
	(0.220)	(0.220)	(1.233)	(0.220)	(0.220)	(1.233)
observations	72296	72293	51099	72296	72293	51099
R-squared	0.30%	0.30%	0.90%	0.30%	0.30%	0.90%

Note: This table reports the multivariate tests of the effects of the relative net order flows measure (RNOF) on the post-announcement cumulative abnormal return. The dependent variable is CAR [2,61], which is defined as the difference between the buy-and-hold return of the announcing firm and that of a size and book-to-market matching portfolio over the long-term window, starting from the next day after the announcement window [0,1] and covering the subsequent 60 trading days [2,61] relative to the announcement window. The relative net order flows measure is computed using order-level data from 1983 to 2005 from ISSM and the TAQ dataset. RNOF = average standardized net order flows of nonannouncing firms – standardized net order flows of the announcing firm over the announcement window. Trade direction is inferred following the standard five-second rule (Lee and Ready, 1991). ESD is the earnings surprise decile. NRANK is the number-of-announcements decile. The control variables include firm's size, book-to-market ratio, earnings persistence, earnings volatility, the number of analysts following the stock, reporting lag, institutional ownership, share turnover, and year, month, and Friday time effects. All coefficients are expressed as percentages. Standard errors are reported in parentheses. * Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level.

Table 5

Liquidity and Order Imbalance

Panel A. Correlation Coefficient between Standardized Order Imbalance Rank and Liquidity Rank

	Liq-Rank (MQI)	Liq-Rank (Amihud)	SOI-Rank
Liq-Rank (MQI)			
p-value			
Liq-Rank (Amihud)	0.512		
p-value	0.00		
SOI-Rank	0.116	0.213	
p-value	0.00	0.00	

Panel B. Regression of Order Imbalance on Liquidity Rank and Earnings Surprise Magnitude

	SOI
Liq-Rank (MQI)	0.014***
t-stat	19.37
|Earnings Surprise|	0.059
t-stat	1.49
Event	0.080***
t-stat	15.91
Event*Liq-Rank	0.020***
t-stat	12.46
Constant	0.500***
t-stat	226.07
R-squared	1.55%
OBS	339541

Note: Stocks are ranked within each event group according to their liquidity (MQI and Amihud measure). A rank of 1 (5) represents the lowest liquidity (that is, it has the lowest MQI or highest Amihud illiquidity measure in that group). Stocks are ranked according to the standardized order flows imbalance within each group to reflect investors' trading interest in a stock relative to other stocks. Standardized order imbalance (SOI) is measured over the two-day event window as the absolute value of the average of standardized daily net order flows, which is the daily net order flows demeaned and normalized by the standard deviation of the net order flows of the stock in the month of announcement. A rank of 1 stands for lowest order imbalance, and a rank of 5 stands for highest order imbalance. The absolute value of earnings surprise reflects the magnitude of the earnings shock. Event is the indicator variable for the stocks of announcing firms. * Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level.

Table 6

Regression of the PEAD on RNOF for Groups with Different Liquidity Ranks

	Share Volume		
	LiqRank=1&2	LiqRank=3	LiqRank=4&5
	CAR[2,61]	CAR[2,61]	CAR[2,61]
ESD	0.468*	0.848**	1.165***
	(0.283)	(0.378)	(0.308)
RNOF	0.293*	0.185	0.220
	(0.174)	(0.217)	(0.174)
Controls and controls interacted with earnings surprise decile	YES	YES	YES
constant	-5.070***	-8.003***	-9.408***
	(1.801)	(2.461)	(2.088)
R-squared	1.40%	1.40%	1.40%
	Dollar Volume		
	LiqRank=1&2	LiqRank=3	LiqRank=4&5
	CAR[2,61]	CAR[2,61]	CAR[2,61]
ESD	0.468*	0.849**	1.165***
	(0.283)	(0.378)	(0.308)
RNOF	0.295*	0.208	0.213
	(0.173)	(0.217)	(0.175)
Controls and controls interacted with earnings surprise decile	YES	YES	YES
constant	-5.070***	-8.006***	-9.408***
	(1.801)	(2.461)	(2.088)
R-squared	1.40%	1.40%	1.40%

Note: This table reports the multivariate tests of the effects of the relative net order flows (RNOF) measure on the postannouncement cumulative abnormal return for event stocks with different relative liquidity in its group. The dependent variable is CAR [2,61], which is defined as the difference between the buy-and-hold return of the announcing firm and that of a size and book-to-market matching portfolio over the long-term window, starting from the next day after the announcement window [0,1] and covering the subsequent 60 trading days [2,61] relative to the announcement window. The RNOF measure is computed using order-level data from 1983 to 2005 from the TAQ dataset. RNOF = average standardized net order flows of nonannouncing firms – standardized net order flows of the announcing firm. Trade direction is inferred following the standard five-second rule (Lee and Ready, 1991). ESD = the earnings surprise decile. The control variables include the number-of-announcements decile interacted with earnings-surprise decile, firm size, book-to-market ratio, earnings persistence, earnings volatility, the number of analysts following the stock, reporting lag, institutional ownership, share turnover, and year, month, and Friday time effects. LiqRank = 1 represents lowest liquidity in the group, and LiqRank = 5 represents the highest liquidity. All coefficients are expressed as percentages. Standard errors are reported in parentheses. * Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level.

Table 7

Regression of the PEAD on RNOF for Groups with Different Earnings Surprises

	Share Volume		
	Earnings Surprise<0	Earnings Surprise=0	Earnings Surprise>0
	CAR[2,61]	CAR[2,61]	CAR[2,61]
ESD	0.862		-0.222
	(0.853)		(0.488)
RNOF	0.197	0.314	0.240*
	(0.181)	(0.266)	(0.140)
Controls and controls interacted with earnings surprise decile	YES	YES	YES
constant	-7.301***	0.640	0.738
	(2.237)	(12.220)	(3.654)
R-squared	1.00%	1.50%	0.80%
	Dollar Volume		
	Earnings Surprise<0	Earnings Surprise=0	Earnings Surprise>0
	CAR[2,61]	CAR[2,61]	CAR[2,61]
ESD	0.862		-0.221
	(0.853)		(0.488)
RNOF	0.186	0.318	0.260*
	(0.181)	(0.265)	(0.139)
Controls and controls interacted with earnings surprise decile	YES	YES	YES
constant	-7.303***	0.636	0.730
	(2.237)	(12.230)	(3.654)
R-squared	1.00%	1.50%	0.80%

Note: This table reports the multivariate tests of the effects of the RNOF measure on the postannouncement cumulative abnormal return for different earnings surprise groups. The dependent variable is CAR [2,61], which is defined as the difference between the buy-and-hold return of the announcing firm and that of a size and book-to- market matching portfolio over the long-term window, starting from the next day after the announcement window [0,1] and covering the subsequent 60 trading days [2,61] relative to the announcement window. The relative net order flows measure is computed using order-level data from 1983 to 2005 from the TAQ dataset. RNOF = average standardized net order flows of nonannouncing firms – standardized net order flows of the announcing firm. Trade direction is inferred following the standard five-second rule (Lee and Ready, 1991). ESD = the earnings surprise decile. The control variables include the number-of-announcements decile interacted with earnings-surprise decile, firm size, book-to-market ratio, earnings persistence, earnings volatility, the number of analysts following the stock, reporting lag, institutional ownership, share turnover, and year, month, and Friday time effects. LiqRank = 1 represents the lowest liquidity in the group and LiqRank = 5 represents the highest liquidity. All coefficients are expressed as percentages. Standard errors are reported in parentheses. * Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level.

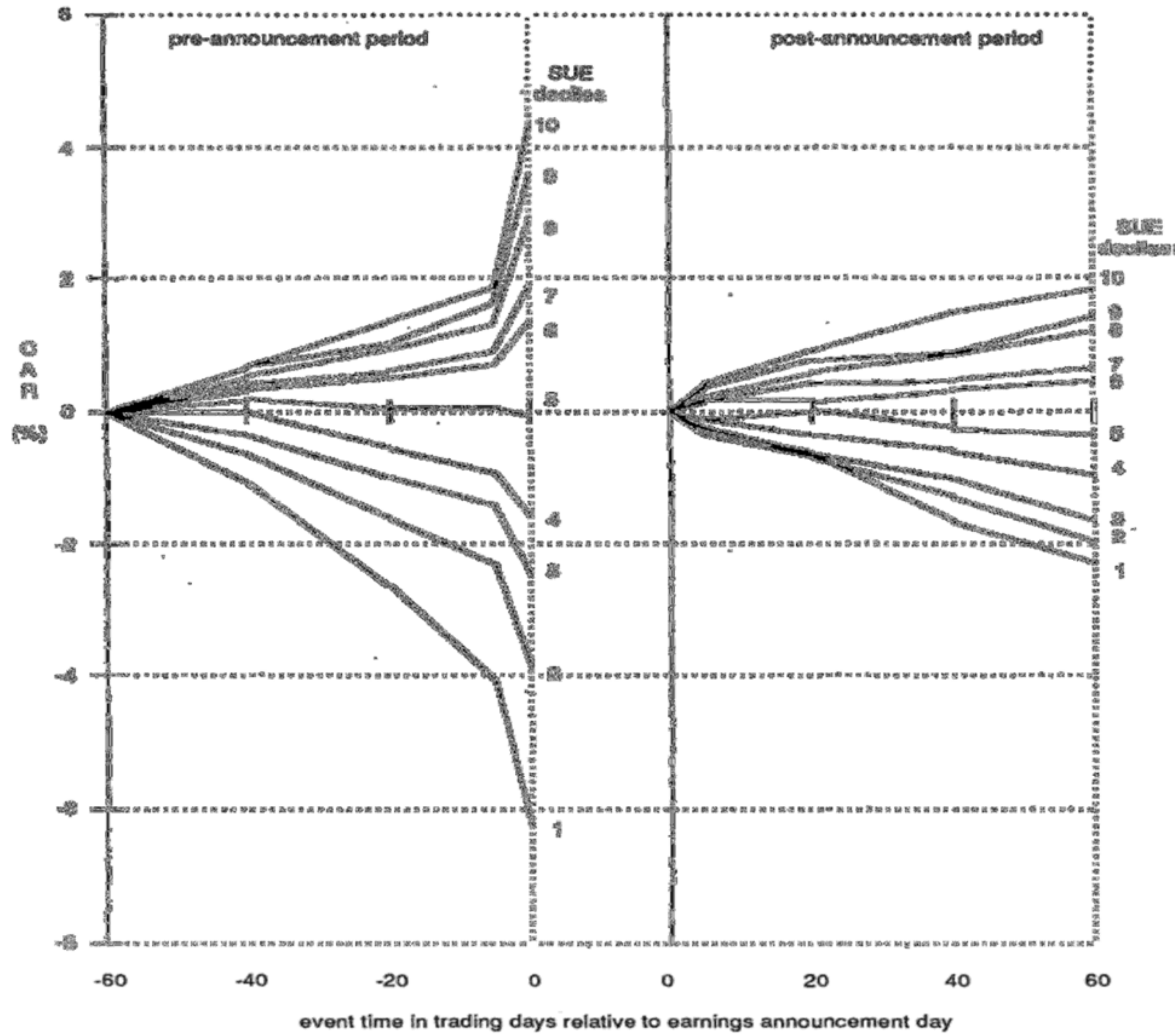
Table 8

Regression of the PEAD on RNOF and Returns

	(1)	(2)
	CAR[2,61]	CAR[2,61]
ESD	0.707***	0.707***
	(0.221)	(0.221)
ACAR[0,1]	-6.840	
	(4.290)	
RCAR[0,1]		-1.804
		(1.279)
RNOF	0.252**	0.252**
	(0.101)	(0.101)
Controls and controls interacted with earnings surprise decile	YES	YES
constant	-7.346***	-7.346***
	(1.215)	(1.215)
Obs	50803	50803
R-squared	0.80%	0.80%

Note: This table reports the effects of stock returns of nonannouncing firms on the postannouncement cumulative abnormal return. The dependent variable is CAR [2,61], which is defined as the difference between the buy-and- hold return of the announcing firm and that of a size and book-to-market matching portfolio over the long-term window, starting from the next day after the announcement window [0,1] and covering the subsequent 60 trading days [2,61] relative to the announcement window. The RNOF measure is computed using order-level data from 1983 to 2005 from the TAQ dataset. RNOF is the average standardized net order flows of nonannouncing firms – standardized net order flows of the announcing firm. Trade direction is inferred following the standard five-second rule (Lee and Ready, 1991). ESD is the earnings surprise decile. ACAR[0,1] is the average cumulative abnormal returns of stocks of nonannouncing firms over the event window. RCAR is the difference between ACAR[0,1] and the event stock CAR[0,1]. The control variables include the number-of-announcements decile interacted with earnings-surprise decile, firm size, book-to-market ratio, earnings persistence, earnings volatility, the number of analysts following the stock, reporting lag, institutional ownership, share turnover, and year, month, and Friday time effects. All coefficients are expressed as percentages. Standard errors are reported in parentheses. * Significance at 10% level. ** Significant at 5% level. *** Significant at 1% level.

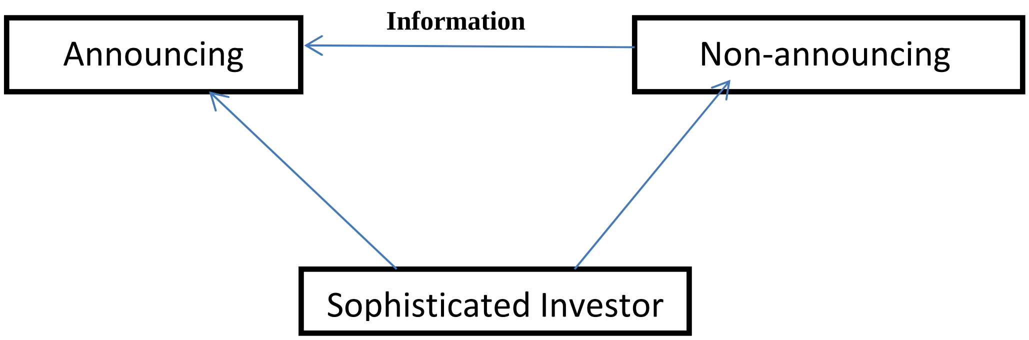
Figure 1
The Post-Earnings-Announcement Drift



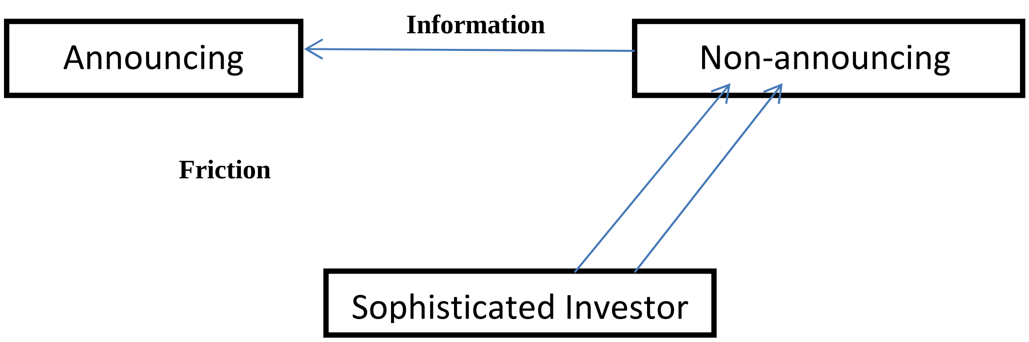
Note: This figure, excerpted from Bernard and Thomas (1989), presents a simple example of the postannouncement abnormal stock returns for each earnings surprise (SUE) decile. Decile = 10 is the best earnings news, and decile = 1 is the worst earnings news.

Figure 2

A Graphical Illustration of the Relation between Non-announcing Stock Trading and the PEAD



Note: Sophisticated investors may choose to invest in both announcing and non-announcing stocks when they receive the earnings announcement.

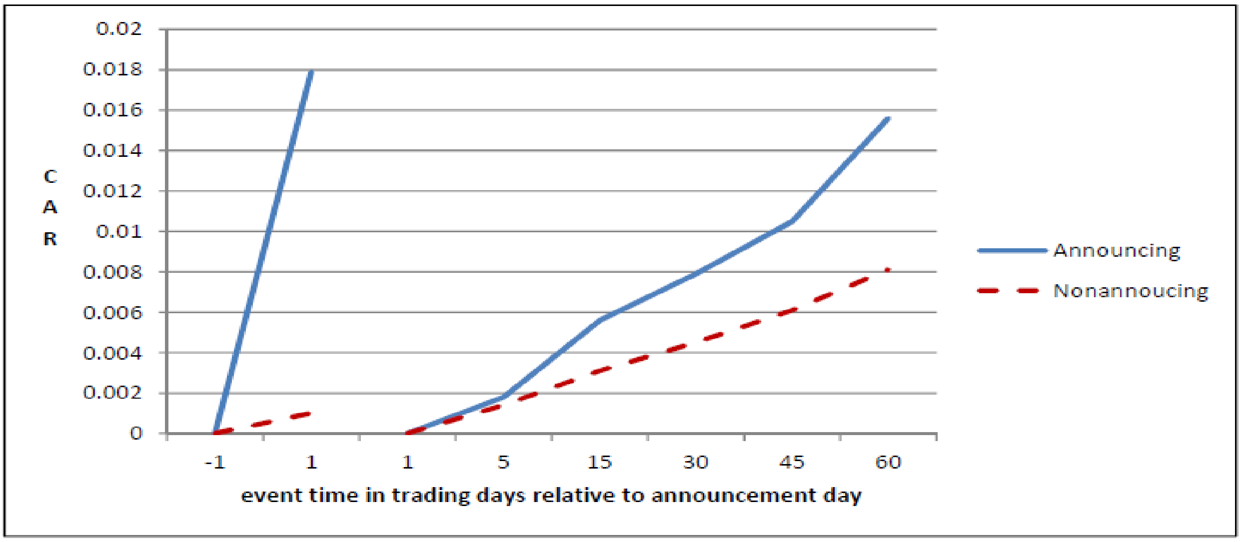


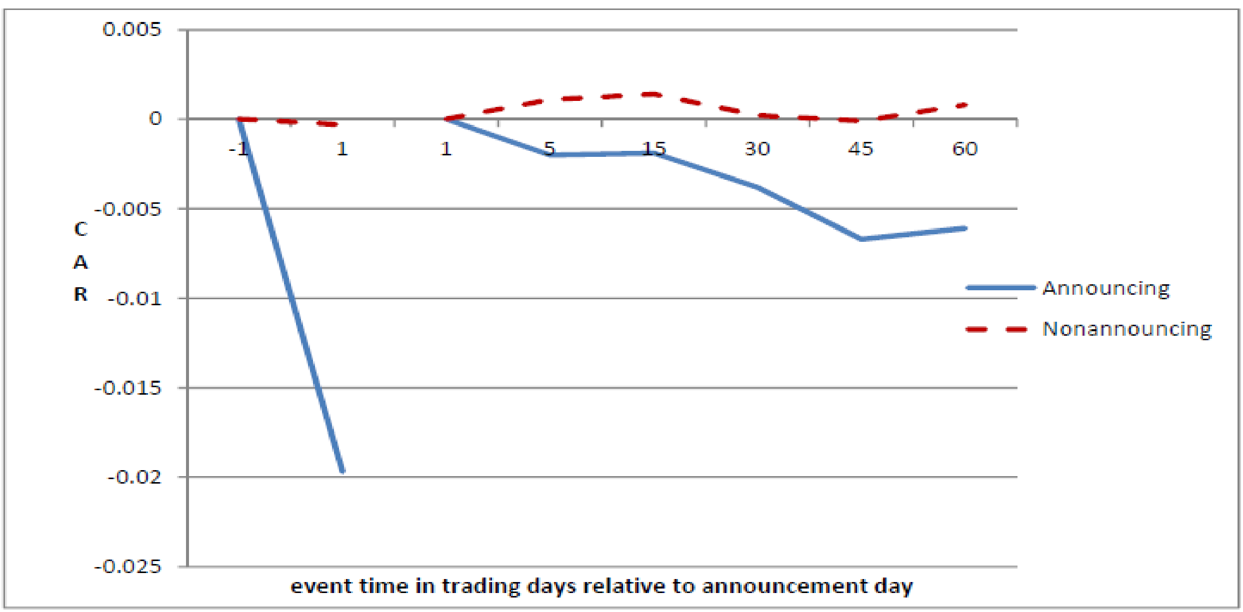
Note: Non-announcing firms' stocks are more attractive to these investors if there is more trading friction in announcing stocks. Trading activities in non-announcing stocks may first reflect information common to the corresponding industry and have predictive power for the price movement of the event stock.

Figure 3

The Drift in the Announcing and Non-announcing Stock Return

Panel A. CARs after Positive Earnings Surprises





Panel B. CARs after Negative Earnings Surprises

Note: This figure plots the average cumulative abnormal returns (CAR) for the announcing firms and the non-announcing firms after good and bad earnings reports. The announcement window CAR and the post-announcement CAR are separated to facilitate analysis.

Chapter 3

Stock Order Flow Commonality: the Role of Liquidity

Abstract

How do stocks with different liquidity move together in response to new information? Are liquid stocks more sensitive to economic news? This paper emphasizes the role played by stock liquidity and argues that stock comovements are closely related to liquidity because stocks with high liquidity would be investors' less costly choice when they want to trade in response to economic information. Constructing a random sample at monthly frequency covering 8 GICS sectors over the period 1993-2005, I show that large and liquid stocks have significantly higher order flow commonality in a panel analysis. From a time series perspective order flow commonality of a group of stocks is more related to aggregate market liquidity rather than the average liquidity in that group. Specifically, an increase in the commonality of medium and large stocks predicts a decrease in aggregate market liquidity one month ahead. Moreover, an increase in commonality among large (small) stocks predicts an increase (decrease) in market uncertainty one or two months ahead. This implies that informed investors prefer to trade large and liquid stocks when the market liquidity is going to decline and the information uncertainty is going to rise.

1. Introduction

Past studies have documented significant common variation in order flows and stock returns. An important issue in finance studies is the factors affecting stock comovements. In general, investors' trading and stock returns tend to move together when these stocks are exposed to common information shocks. Prior research has revealed several factors that have impact on order flow commonality such as stock indexing and program traders. However, as an important factor affecting investors' trading activities, the role of stock liquidity has been ignored in extant literature. The concept of liquidity is closely tied with market friction and trading cost, and therefore should affect investors' decision about where to trade and should have impact on stock comovements.

How is commonality in stock order flow related to liquidity from cross-section and time series perspective? Do we expect to see higher or lower commonality for more liquid stocks? What role does commonality play in relation with market uncertainty? Do changes in commonality lead changes in market liquidity and market uncertainty, or the other way around? These are all questions to be answered in this paper.

Intuitively, we expect higher commonality for more liquid stocks from a cross-sectional perspective. Figure 1 presents a graphical illustration of this hypothetic relation. Assume that there are some informed investors who closely watch the news and have superior analytical skills about the market and some uninformed investors who are just noise traders or liquidity providers. When there is an information shock common to some stocks, the informed investors will trade with uninformed investors accordingly. If there are multiple stocks exposed to the news, they will prefer to trade stocks that have higher liquidity to reduce transaction cost.

Therefore stronger commonality in order flows should be observed among more liquid stocks. For example, suppose there is technology shock hitting a whole industry. Informed investors react to the new information fast by trading related stocks. Facing different liquidity levels of stocks, they choose stocks that are costless and easy to trade, which lead to commonality in order flows among stocks with higher liquidity. Illiquid stocks in that industry, may react to the news as well, but reactions of different stocks may exhibit disperse and nonsynchronous pattern (for instance, illiquid stocks may just update quotes without any trades) since investors trading is impeded by illiquidity, which makes it difficult for researchers to observe commonality in order flows among these stocks.

My study endeavors to build a link between order flow commonality and stock liquidity. I first do a panel analysis of the relation between commonality and liquidity and then investigate time series interaction of these two variables. Specifically I examine commonality of order flows for stocks with different market capitalization in eight out of the ten sectors defined by the Global Industry Classification Standard (GICS) and construct a panel data sample over the period 1993-2005. Consistent with my hypothesis, strong and positive effect of liquidity on order flow commonality is found in the panel data after controlling for a set of relevant variables. Then an event study of federal tax cut announcements shows that comovements of large and liquid stocks react more sensitively to economic news.

Regrouping stocks into small, medium and large size groups, I investigate the time-series relationship between liquidity and order flow commonality for small, medium and large stocks respectively. No significant Granger-causality between these two variables is found for all size groups. Changes in commonality and liquidity are primarily explained by their own lags. Hypothesizing order flow commonality of each group is related to the aggregate market liquidity, I further examine the interaction between commonality and the aggregate market liquidity. I find that an increase in commonality among medium and large stocks predicts a decrease in the aggregate market liquidity level. Lastly I examine whether there is interaction between market uncertainty and commonality. The results suggest that commonality among large stocks is contemporaneously positively correlated with information uncertainty measured by VIX. Moreover, an increase in commonality of large (small) stocks predicts an increase (decrease) in market uncertainty. This is consistent with the fact that investors tend to trade more liquid stocks when the expected uncertainty is high and the expected market liquidity is low.

My study points out the important role played by stock liquidity and complements extant literature studying driving factors of order flow commonality. To my knowledge, it is the first to study the variation of order flow commonality over time and bring up an interesting question about how to interpret the information content in order flow commonality. It shows that commonality changes lead the market liquidity and market uncertainty changes, suggesting that commonality contains useful information for future movements of other economic variables.

The rest of the paper proceeds as follows. Section 2 discusses how this paper is related to prior research. Section 3 discusses data and sample construction. Section 4 conducts a panel data analysis of the relation between order flow commonality and stock liquidity. Section 5 investigates the time series relation between liquidity and commonality for small, medium and large size groups as well as the relation between commonality and aggregate market liquidity. Section 6 studies the time series interaction between commonality and market uncertainty. Section 7 provides concluding remarks.

2. Literature Review

Recently, a growing body of research has focused on utilizing the order-level trading data to discover information that is not conveyed in low-frequency data. Stock order flows have been shown to contain important information beyond low-frequency data. For example, Chordia and Subrahmanyam (2004), and Chordia, Roll and Subrahmanyam (2005) find that net order flows (with individual orders signed as in Lee and Ready (1991) algorithm) are positively related to contemporaneous and future returns at daily and intra-day interval. Looking at longer horizons of between six months to two years, Hvidkjaer (2008) find that future returns are negatively related to net order flows of small trade. In particular, the stocks that small investors prefer to hold underperform those sold by them on average, which suggests that small investors are more likely to have inaccurate valuation of stocks and are outplayed by large investors. In a recent paper, Beber, Brandt and Kavajecz (2011) examine the relation between sector order flows and the macro economy. They find that unexpected net order flows into the materials/manufacturing sector forecast economic expansions, while net order flows into the utility and financial sectors, indicate economic recessions.

Since intra-day data contain important information about the stock market and macro economy, researchers are interested in whether there is cross sectional common variation in order flows. Hasbrouck and Seppi (2001) examine the commonality in order flows, returns and liquidity for Dow 30 stocks. They find a significant common factor in stock order flows and find that the first principal factor explains approximately two-thirds of the common factor in returns. Harford and Kaul (2005) examine commonality in order flows for S&P stocks and non-S&P stocks. They find stronger commonality in S&P stocks and conclude that stock indexing is a primary factor that drives commonality in order flows. Hughen and McDonald (2006) examine commonality in order imbalances across different types of securities and find commonality present in small and medium trades, but not in large trades, across portfolios of small stocks, large stocks, and closed-end funds. Order imbalances from larger size trades provide more information relevant to stock returns of other portfolios. It is still an interesting and open question what the economic factor that drives the commonality in order flows is and how that factor is related to stock market performance.

There are several prevailing explanations for the commonality in order flows. In a nutshell, these potential explanations argue that commonality can be attributed to some common characteristics either in investors or in securities, which expose a set of stocks to a common information shock. Therefore these explanations can be classified into two categories.

The first category of explanations is based on similarities shared by investors. For example, Corwin and Lipson (2011) argue that commonality is driven by program trading or institutional traders. Using a dataset consisting of electronic order flow data for a sample of NYSE-listed stocks, they find that commonality in order flows is primarily driven by the correlated trading decisions of professional traders but not retail traders. Another example is Lee, Shleifer, and Thaler (1991), who argue that habitat investing by subsets of investors may result in commonality in the returns of the securities held by them. They find evidence of commonality in the prices of closed-end funds which are held primarily by small individual investors. Also they find that these funds co-move with small stocks.

The second category of explanations relies on the similarities among securities. Consider Harford and Kaul (2005) discussed above as an example. Barberis and Shleifer (2003) develop a model where investors categorize assets into different styles. They argue that assets that share common characteristics, which can be based in law, in markets or in fundamentals, can be categorized into a style or class by investors. Investors move funds among the stock styles depending on their relative performance, which drives commonality in order flows and returns in the stocks of the same style. Another potential explanation under this category involves differences in the speed of information diffusion. Barberis, Shleifer and Wurgler (2005) provide evidence for this explanation. They show that a stock's beta with the S&P portfolio goes up after addition to S&P and goes down after deletion from S&P stocks. They further decompose the change in beta around stock addition and deletion into a component due to information diffusion, and a component likely due to habitat investing and style investing and find that the information diffusion story account for a portion of commonality across stocks.

The importance of stock liquidity in explaining comovements is very much related to the information diffusion story since liquidity can affect the speed at which stocks reflect economic information. Higher liquidity suggests lower transaction cost and more efficiency in reflecting information and thus leads to higher commonality among liquid stocks. However, it is worth noting that the role of liquidity is not limited to information diffusion story. In fact, liquidity is a necessary condition for observing commonality in order flows in all of the explanations above. For example, past studies suggest that stock indexing, industry and program trading may be potential factors driving commonality. These stocks are typically more liquid stocks and therefore it is not surprising that these stocks have higher common movements in response to newly released economic information when investors prefer to trade in these stocks. I will show shortly that the liquidity has stronger and more robust effect than indexing and intuitional trading.

3. Data and Empirical Methods

To study the effect of liquidity on order flow commonality, I first categorize stocks by the sectors they belong to defined by the Global Industry Classification Standard (GICS). GICS is an industry taxonomy developed by MSCI and Standard & Poor's (S&P). There are in total 10 sectors, 24 industry groups, 68 industries and 154 sub-industries. Then I divide each sector into size terciles (small, medium and large with cutoff points defined within each sector) by their market capitalization. Then I randomly select 30 stocks from each tercile of these sectors to form a sector-size group. Sectors with code 50 (telecommunication services) and 55 (utilities) are excluded from the sample due to insufficient observations. Hence there are 24 groups over the period 1993-2005. The members of stocks for each group are renewed every 5 years because firm size changes over time.

Transaction and quote data are obtained from Consolidated Trades and Consolidated Quotes datasets respectively from the Trade and Quote (TAQ). To calculate order flow, each trade is signed following the 5-second rule proposed by Lee and Ready (1991).[7] Transactions that occur at prices above the midpoint are classified as buys (positive sign) while those that occur at prices below the midpoint are sells (negative signs). Trades occurring at the midpoint are excluded from the sample. Stock order flows are defined either as the signed share volume or dollar volume.[8] Transactions occurring outside the trading hours (9:30AM to 4:00PM) are excluded from the sample. Following Hasbrouck and Seppi (2001), I cumulate order flows for each 15-minute interval covering 9:30-9:45, 9:45-10:00, …, 15:45-16:00 for a total of 26 intervals in a trading day. As Hasbrouck and Seppi (2001) point out, the use of 15-minute

[7] The result is robust to different time adjustments such as zero, one and three second adjustments.
[8] These two measures are very highly correlated after standardization discussed below and thus yields the same result.

intervals represents "a compromise between, on the one hand, needing to look at correlations in contemporaneous order flows across stocks and, on the other, seeking to minimize return/order flow simultaneity problems." In other words, a 15-minute resolution represents a balance between minimizing the risk of confounding non-contemporaneous order flows and maximizing the time allowed for investors to react to some common information shock. To construct a panel data sample, commonality in order flows is examined at monthly frequency. Therefore I have 3744 observations over 13 years. To remove the time-of-day effects documented in Wood et al. (1985), order flows in an interval are standardized by their monthly mean and standard deviation for that interval. For example, let "OF" denote order flows measure and $OF_{i,m,d,k}$ denote the order flows for stock i for 15-minute interval k on day d in month m. The standardized order flows measure is computed as $OF^{*}_{i,m,d,k} = (OF_{i,m,d,k} - \mu_{i,m,k})/\sigma_{i,m,k}$ where $\mu_{i,m,k}$ and $\sigma_{i,m,k}$ are the mean and standard deviation for order flows of stock i and interval k, estimated across days in month m. Unlike prior research, I do not classify trades into small, medium and large groups to examine trade size effects. The reason is that at the monthly frequency, there are often no observations of large transactions for a given stock, which will reduce the number of stocks available in the principal component analysis and contaminate the results. The principal component analysis (PCA) models a set of factors as mutually orthogonal linear combinations of standardized stock order flows. This procedure allows me to extract a set of factors that collectively maximize the overall explanatory power. The explanatory power of a principal component is measured by the proportion of total variation of order flows explained by that component. Since the order flows variable has been standardized to unit variance, the total variation of a group in a month is simply the number of stocks, 30.

Stock liquidity is an elusive concept. There are many different measures of stock liquidity employed in the literature. In this paper, I select the widely used Amihud (2002) price impact measure as a proxy for transaction cost. This measure uses low-frequency daily data to capture the "daily price response associated with one dollar of trading volume.'' Goyenko, Holden, and Trzcinka (2009) compare multiple different proxies of liquidity based on low-frequency data and argue that "the widely used Amihud (2002) measure is a good proxy for price impact." Specifically, the monthly stock illiquidity measure for stock *i* is computed as the average of the daily price impact

$$illq_{i,m} = \frac{1}{T}\sum_{t=1}^{T}\frac{|r_{i,t}|}{Dvol_{i,t}}$$

where T is the number of days in month m, $r_{i,t}$ is the return for stock i on day t and $Dvol_{i,t}$ is the dollar volume traded for stock *i* on day *t*. Then the liquidity of a group is simply the equal-weighted average of the individual stock liquidity in that market.

Another alternative measure for liquidity is bid-ask spread. Bid-ask spread is based on high-frequency trading data. I use both quoted spread and effective spread, which are respectively defined as

$$Qspread_{i,m} = \frac{1}{n_{i,m}}\sum_{j=1}^{n}\frac{Ask_{i,j} - Bid_{i,j}}{mp_{i,j}}$$

and

$$Espread_{i,m} = \frac{1}{n_{i,m}}\sum_{j=1}^{n}\frac{|p_{i,j} - mp_{i,j}|}{mp_{i,j}}$$

where $mp_{i,j}$ =($Ask_{i,j} - Bid_{i,j}$)/2; $p_{i,j}$ is the transaction price for the jth trade of asset i in month m; $n_{i,m}$ is the number of eligible trades of stock i in month m; ; $Ask_{i,j}$ and $Bid_{i,j}$ are the ask and bid quotes prevailing at the time of the jth trade of asset i in month m. The liquidity of a group is defined the same way as the equal-weighted average of the individual stocks in that market. The results based on these spread measures are not tabulated but summarized in the footnote in the next section.

4. Panel Analysis

In this section I examine the commonality in order flows by estimating the principal components for different sector-size groups. The results show that in general, groups composed of large stocks have higher order flow commonality. GICS sectors of materials (code 15) and information technology (code 45) exhibit higher commonality while GICS sectors of materials (code 20) and consumer discretionary (code 20) have lower commonality despite that the difference is not sizeable. Monthly commonality is then put into a panel regression to study the relation between stock liquidity and order flow commonality. Section 4.3 conducts an event study on the effects of federal tax cuts on stock order flow commonality.

4.1 Principal Components

Table 1 reports the estimated commonality using PCA. For each sector-size group, the time-series average of each of the first three eigenvalues is presented. There is a clear pattern that large stocks tend to have a higher first eigenvalue. The first eigenvalue ranges from 1.62 to 2.06 for large stocks while and 1.33 to 1.63 for small stocks, implying that at most about 6.8% (2.06/30) of variance is explained by the first common factor for large stocks, while 5.4% (1.63/30) is explained for small stocks. Table 2 presents the cumulative variation of order flows explained by the first three factors. Again, large stocks have the highest explained proportion of total variation. The materials sector (code 15) and information technology sector (code 45) exhibit the highest commonality while the industrial sector (code 20) and consumer discretionary sector (code 25) show the lowest commonality.

It is an interesting question how order flow commonality varies over time and how the variation differs for different sectors. Figure 2 depicts the time variation of the first eigenvalue of order flow commonality by GICS sectors. The figure delivers several interesting observations. Firstly, there is a clear relation between commonality and firm size. For all sectors, in general medium and large stocks have higher order flow commonality than small stocks and most of the cases large stocks have highest order flow commonality. The size effect on commonality is strong and clear, which is suggestive about the role played by stock liquidity since size is usually highly correlated with liquidity and large stocks are more liquid stocks. It appears that investors prefer to trade in large stocks when some new information about the economy or a specific sector arrives.

A second observation is that commonality among medium and large stocks is more volatile over time. In each of these eight pictures, commonality exhibits a number of ups and downs for large stocks while appears to be rather stable for small stocks. Higher volatility indicates that large stocks react more frequently and rapidly to information shocks, which is consistent with my hypothesis that high liquidity results in more sensitive stock comovements in response to economic news.

Thirdly, order flow commonality across sectors shares some common pattern over time, suggesting that there is some underlying economic variable that is correlated with commonality in all sectors. More specifically, medium and large stocks appear to have stronger co-movements, suggesting these stocks may be more sensitive to that economic factor. It can be seen that for all sectors, order flow commonality starts off at low levels in our sample then starts to pick up before 1997. Then it evolves at low levels for post-1997 period until somewhere around early 2003. Then it goes up quickly after the burst of dot come bubble. It would be useful if we examine how commonality changes at points of some major events. I mark the breakout of the Asian Financial Crisis in 1997 with a straight vertical line. October 1997 is approximately the time when the U.S stock market was hit by the crisis. Note that there is a sharp decline in commonality in October 1997 in all sectors, suggesting commonality is affected by some underlying market wide economic variable. Since this paper investigates the relation between stock liquidity and order flow commonality, I conjecture that aggregate market liquidity is a relevant variable. Table 3 reports the contemporaneous correlation coefficients of order flow commonality among large stocks and monthly aggregate market liquidity. The monthly aggregate market liquidity is measured as the equal-weighted Amihud price impact illiquidity measure across all stocks available on CRSP dataset. Consistent with my hypothesis, most of these sectors have a negative correlation coefficient, implying these sectors have higher commonality when market liquidity is high. In particular, the energy, materials, industrial and financial sectors have significant and negative coefficients, suggesting that investors' trading is more sensitive to aggregate market liquidity. When market liquidity dries up, these sectors are more likely to experience decreased informed trading activities and commonality in order flows. The consumer staples, consumer discretionary, health care and information technology sectors do not have statistically significant correlation with market liquidity, suggesting that commonality in these sectors is more likely correlated with some sector-specific factor or other underlying macroeconomic factors. It is worth noting that these correlation coefficients are simply computed from two crude contemporaneous time series. So far I have not done any tests for unit roots or time series stationarity. More complex and advanced time series analysis will be conducted in the next section.

Fourthly, order flow commonality of stocks with different sizes exhibits co-variation within each sector, implying that commonality is correlated with some sector or industry-specific factor. For example, the figure shows substantial co-movements of monthly order flow commonality across size groups within each sector. A more vivid illustration to the sector-wide co-movement is the “Dot Com Bubble” period, which is shaded in Figure 2. During the “Dot Com Bubble,” all sectors except energy and information technology have relatively low order flow commonality compared to in other periods in history. This is not surprising since during the bubble periods, investors are fanatical about IT stocks. They watch economic news closely and are very sensitive to any information related to IT industry in the market. Information shocks will be quickly reflected in stock order flows in information technology industries while stocks from other industries do not have much informed trading.

The fifth observation is that there is a significant time trend in order flow commonality. The average commonality level in the second half of our sample period is significantly higher than the level earlier in time, which may be attributed to faster information dissemination via modern development in mass media technology and more comovements in computer-based stock trading.

Lastly, it appears that on average, information technology stocks tend to co-move more than stocks of other sectors. This may either suggest that compared to other industries, IT stocks are more likely to be exposed to economic information shock or that investors' attention and trading is more focused on these stocks. In fact, there is not very significant difference in commonality across sectors, suggesting sector or industry, among factors driving commonality, does not contribute very much to it. This is consistent with Harford and Kaul (2005), who point out that in their sample, industry effects on commonality exist but are not very strong and that the indexing effect is dominant. Similarly, the sector effect is not strong in my sample, while the size effect is dominant.

To further investigate how liquidity is related to order flow commonality, a panel data analysis is conducted in the following section.

4.2 Panel Regressions

The relation between stock liquidity and order flow commonality can be estimated by a panel regression. Recall that we have a panel sample tracking 24 groups over 13 years and the assumption of independent sampling is violated. Thus when estimating the regression, we need to take care of both cross-sectional and serial correlation to obtain robust standard errors.

Specifically, I estimate the following regression:

$$com_{i,t} = \alpha + \beta groupliq_{i,t} + \gamma X_{i,t} + Fix_{time} + Fix_i + \varepsilon_{i,t} \tag{1}$$

where $com_{i,t}$ is the commonality of order flows for group *i* at month *t*, $groupliq_{i,t}$ is the decile number ranked within my sample by the group liquidity computed as the equal weighted average of its members' liquidity, $X_{i,t}$ is a vector of control variables including aggregate market liquidity, the number of S&P indexing stocks and the average institutional holding of stocks and in group *i* at month t, Fix_{time} and Fix_i are time and sector fixed effects respectively and $\varepsilon_{i,t}$ is the error term. In this regression, I use the sum of the first three eigenvalues to measure commonality.[9] I use the decile number (higher number represent higher liquidity) of group liquidity instead of liquidity itself to allow for a non-linear relation. Using contemporaneous liquidity measure might create endogeneity and simultaneous causation problem, which will lead to estimation bias. Therefore I also sort the stocks into decile based on lagged liquidity and find that the same results hold. The inclusion of the number of index stocks and the average institutional holdings is supported by prior studies. Because my sample is "narrow and long," which means that it has a large number of observations on time dimension and relatively less cross sectional observations, it is more time-series-cross-sectional (TSCS) than a regular panel with more cross-sectional observations. The TSCS sample is more widely used in political science when there is a small number of spatial units (for example, states and nations) and large number of time units. We need to be precautious before estimating such a model since the traditional OLS standard error assumption will be violated most of the time. There may be at least three types of violations. First, errors tend to be heteroskedastic. As noted above, Figure 2 provides some evidence that large stock exhibits higher volatility in order flow commonality. Second, errors are likely to be serially correlated (temporal dependence) over time for each group. This is an important issue when facing a sample with more time units than spatial units. Third, errors may be contemporaneously correlated across groups. This issue is relevant since our analysis in section 4.1 has shown that there may be some underlying economic variables driving commonality for all groups and thus errors are likely to be correlated. To fix these problems, I employ the Beck and Katz (1995) (BK) method, which is the most widely used method in estimating TSCS sample. They show that their estimator performs well when the

[9] The use of the sum of the first 3 eigenvalues leads to very similar results as if only the first eigenvalue is used.

panel's cross-sectional dimension, N is much smaller than its time dimension T. BK assumes the following structure of error covariance and evolvement.

$$\mathrm{E}(\varepsilon_{it}^2) = \sigma_i^2 \tag{1.1}$$

$$\mathrm{E}(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij} \tag{1.2}$$

$$\varepsilon_{it} = \rho_i \varepsilon_{it-1} + v_{it} \tag{1.3}$$

This specification allows for heteroskedasticity across panels and contemporaneous correlation between panels. It assumes an AR(1) process for the residuals to estimate the serial correlation in errors over time.

The BK method involves running an OLS regression first to obtain the residuals. Then it eliminates the autocorrelation in residuals by fitting the AR(1) model and use these residuals to estimate the panel covariance matrix. The "panel corrected standard errors" (PCSE) are then applied to estimate the model and derive statistical inference.

Table 4 presents the BK estimation results under different model specifications. Consistent with my hypothesis that stock groups with higher liquidity tend to exhibit higher order flow commonality, all of these regressions show positive and very significant effects at 1% significance level of the group liquidity (the variable groupliq) on order flow commonality even after control variables are introduced.[10] Regression (1) and (2) are estimated using strongly balanced data of a full sample (3744 observations). Regress (2) includes aggregate market liquidity and the number of S&P indexing stocks. The effect of market liquidity (the variable markliq), however, is not significant after controlling for group liquidity. The coefficients on

[10] Using contemporaneous liquidity measure may cause endogeneity problem. I use the historical liquidity computed as the average monthly liquidity in the past 3 months and get very similar results.

number of indexing stocks (index) in a group are positive and significant, suggesting that consistent with Harford and Kaul (2005), commonality is partially driven by stock indexing. However, I find the incremental explanatory power of these control variables is low by comparing R-squared across regressions. Including only the group liquidity yields a R-squared about 20%. Adding market liquidity and the number of index stocks lead to a very small increase in R-squared. It can be argued that although stock indexing is related to commonality, it does not provide substantial incremental explanatory power once we account for the stock liquidity. Not surprisingly, lagged commonality contributes substantially to R-squared in regressions (3), (4) and (5).

Regression (3) introduces to the right hand side the lagged commonality (COM_{it-1}) and makes the sample size 3720. The inclusion of the lagged commonality is meant to control for the time series evolvement in order flow commonality itself and prevent spurious relations between explanatory variables and commonality. Time and sector-size group fixed effect variables are incorporated in regression (4) and (5) to control for common trend in time and groups. Also I add to regression (4) the decile number of firm size and the average number of business segments in that group as control variables. These two variables are introduced to disentangle the effect of firm size and liquidity. The idea is that liquidity is usually highly correlated with firm size and larger firms are more likely to be exposed to common information shock (or in other words, individual shocks to large firms are more likely to be correlated), which also leads to high commonality among large stocks. The results in Table 4 confirm this hypothesis. I find the average number of firm segments in a group has a very significant and positive effect on commonality. Size and index effects, after controlling for liquidity and # of segments are not as strong though. However, the estimated coefficient on group liquidity remains positive and significant, suggesting liquidity, aside from market capitalization and number of segments, has separate effect on commonality. To confirm the finding that the result is not completely driven by correlated shocks to large firms consisting of multiple segments, I focus on stand-alone firms only and still find higher commonality for large and liquid stocks over the whole sample period. The results are graphed in Figure 3.

Regression (5) has about half of the sample due to data availability on institutional ownership. I find that the average institutional ownership in a group has insignificant effect on commonality after controlling for size and # of segments (the coefficient is significant if we do not control these two), which implies that prior findings of Corwin and Lipson (2011) that program trades and institutional trades are primary driving factor of order flow commonality may be attributed to firm's size and segment number.

In this section, I have demonstrated the importance of liquidity of a group of stocks for the order flow commonality of its member stocks. I show that large size stocks have higher commonality, and commonality for different sectors is negatively correlated with aggregate market liquidity. Using the Beck and Katz (2005) panel corrected standard error estimation method to correct for error autocorrelation, contemporaneous cross-sectional correlation and heteroskedasticity, I find strong and robust effects of group liquidity on order flow commonality.[11] The effects of stock indexing and institutional ownership, after controlling for size and average segment number, are less important.

4.3 An Event Study-Nonfarm Payroll Employment

[11] Using quoted and effective spreads generates very similar results. However, it appears that the effect of effective spread is not robust once the firm size introduced as control variable. The effect of quoted spread is robust.

To see how commonality in large stocks and small stocks react differently to stock market news, it is useful to conduct an event study. Recall that my hypothesis is that when there is common information shock hitting the market, large and more liquid stocks are supposed to reflect the new information more quickly and significantly with stronger order flow co-movements. This event study requires selection of appropriate macroeconomic events. Qualified events are supposed to satisfy the following criteria.

1. Recall my measure of order flow commonality is at the monthly frequency. Therefore the economic events cannot occur too frequently otherwise there is no benchmark for us to observe commonality change.

2. The events should have impact on the whole stock market almost evenly. In particular, the exposure of stocks to the events should be biased towards neither liquid stocks nor illiquid stocks.

3. The events should have clear impact on commonality. Events that create uncertain impact on order flow commonality will make it difficult to observe the effect.

4. The events should deliver information "shocks" to the market. Any information that has already been incorporated into stock market will not induce new trading.

Based on these criteria, the non-farm payroll announcement is a good candidate for the study. Non-farm payroll employment is an influential economic indicator released by the United States Department of Labor on the third Friday after the conclusion of the reference week. It reveals unemployment rate and average hourly earnings of the overall workforce excluding farm workers, private household employees, non-profit organization employees and government employees and thus provides tradable economic information for the stock market. Investors will form their perspective about the economy and trade accordingly after the announcements. My hypothesis is that order flow commonality will increase in the post-announcement period and the increase is expected to be significantly greater for large and liquid stocks. To test this hypothesis, order flow commonality is measured at weekly frequency. For each group of stocks defined as in Section 4.1, I compute the order flow commonality for each week and compare the values of the two weeks after announcement to those of the two weeks before the announcement. Table 5 provides a summary of my analysis. The results are consistent with my hypothesis. I find order flows commonality of large stocks have the most significant increase compared to both of long-run mean and pre-event two weeks average by about 0.54. The commonality among small stocks experiences a negative change after event and is lower than their long-run means. The post-event commonality of order flows of medium stocks has slightly increased compared to its long run mean and has significantly increased compared to its pre-event 2 week average.

In another within-sector event study, I investigate how order flow commonality reacts to news on oil price changes between 1993 and 2005. The results in an unreported table confirm my findings in Table 5. These event studies provide evidence for my argument that large and liquid stocks respond to information innovation more quickly and more effectively and thus enhances our understanding about how the stock market reacts to new information and how information is incorporated into trading and stock price.

5. Time Series Analysis

The panel analysis has shown that order flow commonality is closely related to the group liquidity. Rigorous time series analysis is needed to better understand the relation between liquidity and commonality. Before reaching any reliable conclusion, one should select a proper model to estimate the relation between commonality and liquidity. This section is structured with three parts. In the first part the presence of unit roots is carefully discussed and tested. In the second part, I examine how commonality of order flows in a group is related to the average liquidity of stocks in that group. The Granger-Causality between commonality and liquidity is estimated using both the Vector Autoregression (VAR) model and the vector error correction (VEC) model. In the third part, I investigate the relation between commonality and aggregate market liquidity by repeating the same procedures.

5.1 A Discussion of Unit Root Test

It has long been a controversial issue whether economists should trust unit root tests to determine if a time series is stationary or contains unit roots. At the heart of the controversy is the fact that the unit root processes cannot be "distinguished" from stationary ones in a finite sample, as Cochrane (1991) points out. In fact, it is often the case that unit root tests yield conclusions that counter economic theories. For example, a unit root test for the three month T-bill rate fails to reject the hypothesis of the presence of unit roots and suggests a model specification using differenced data. However, economic intuition tells us interest rate is more likely to be stationary. Interest rate is not expected to "explode" and go to infinity even if we allow for infinite time horizon. Cochrane (1991) provides an example to illustrate this point: "Interest rates were about 6% in ancient Babylon; they are about 6% now." Hence the results of popular unit root tests may not easily translate to a time series model. Since no one really knows the true underlying data generating process, how do we decide which economic specifications to use for given data? Here is my answer: when selecting economic models, we should follow economic theories and use appropriate specification for estimation. But specifications based on unit root tests should be taken into account as well when we examine relations between variables over a short horizon. Cochrane (1991) shows in a Monte-Carlo experiment that models based on differenced data perform better in predicting future movements one period ahead even if the true data generating process is stationary. In the context of my study, economic theories suggest that order flow commonality and stock liquidity are both stationary variables but we are interested in the interaction between these two series over a short horizon. Thus I first test for unit roots in commonality and stock liquidity and then estimate the relation using both VAR model and VEC model. I show that the results generated by these two approaches are similar.

5.2 Commonality and Group Liquidity

To facilitate my analysis, I reclassify stocks into 3 groups that comprise small, medium and large stocks respectively. Grouping the stocks by size dimension only is justified by the result obtained from last section that the size effect dominates the sector effect. Having 3 groups rather than 24 makes my analysis more concise and tractable as well as allows me to examine the results across different size groups. Each sector contributes an equal number of stocks to the new size groups and each of the three size groups consists of 30 x 8=240 stocks that are from 8 GICS sectors. For example, the new group of small stocks is composed by collecting all the 30 small stocks from each sector. Monthly order flow commonality among 240 stocks for each size group is computed in the same way as in last section. Figure 4 plots the time series of order flow commonality proxied by the first eigenvalue for the 3 size groups. The findings documented in last section are confirmed in this figure. Medium and large stocks tend to have higher commonality than small stocks and commonality is more volatile for medium and large stocks. Moreover, the calculated pairwise correlation coefficients suggest that medium and large stocks have a stronger tendency to move together with a coefficient of 0.58. In contrast, the correlation coefficients between small and medium, and small and large stocks are 0.23 and 0.20

respectively. Comparing these correlation coefficients sheds some light on my time series analysis. If there is a macro-level economic variable (for instance, market liquidity) that is related to market-wide time variation of order flow commonality, then medium and large stocks should exhibit higher sensitivity to this factor. In fact, in section 3.2 I confirm this hypothesis and show that relation between market liquidity and commonality is significant for medium and large stocks while insignificant for small stocks. Group liquidity is defined the equal weighted average of individual liquidity as in last section.

To begin, I apply the augmented Dickey-Fuller (DF) test to check if there are unit roots in these two time series for each size group. The augmented DF test includes more than one lags in the test and therefore I need to choose the number of lags included. I follow Schwert's (1989) rule of thumb: # of lags=$12(\frac{T}{100})^{0.25}$. Since T=156 in my sample, I include 13 lags in the DF test. Table 5 reports the statistics of the DF fuller test for these two series in each group. I also conduct DF tests with a time trend. The untabulated test results are similar to those reported in Table 5. All of these DF tests have failed to reject the null hypothesis that there is a unit root present suggesting that these time series are non-stationary.

As discussed in Section 5.1, I first estimate a VAR model. The estimation results are reported in Table 7. It can be observed that in general, the Granger-Causality between commonality and group liquidity is weak. Both lagged commonality and lagged group liquidity do not provide significant explanatory power for the level change in the other variable. Table 8 reports the estimates of the orthogonalized impulse response functions. The orthogonalized impulse response functions can be interpreted as the impact of an unexpected shock in one variable on another variable assuming no other shocks. Cholesky decomposition is applied to pin down the transform matrix. The order of variables is selected based on the VAR regression results. Consistent with the results in Table 7, the impulse response functions suggest weak suggest weak response of one variable to another.

We need to take unit roots into account and examine some short-run properties of these two series. To do this, I test for a cointegration relationship between these two series. Briefly speaking, if unit roots of different time series can be removed by forming a linear combination of these non-stationary series, we say that these series are cointegrated and the linear combination is recognized as the cointegration relation(s) or equation(s). Although usually a unit root can be removed by simply taking the first-order difference, one should be careful with this procedure. The reason is that failing to incorporate the cointegration relation will lead to misspecified VAR models with omitted variables when examining Granger-Causality between variables. The most widely used test for cointegration relation is Johansen's (1988) test, which is based on the VAR model. When the VAR model is transformed into the vector error correction representation, the rank of the coefficient matrix on the lagged variables indicates the number of cointegration equations and there are at most N-1 cointegration equations for N variables. The basis of Johansen's (1998) test is that since the coefficient matrix should be short ranked when there exist cointegration relations, the model should not suffer a loss of fit when a restriction on matrix rank is imposed. Using maximum likelihood estimation, the test has the null hypothesis that the rank of the matrix is less than or equal to a certain number. In our case there are two time series variables, indicating that there is at most 1 cointegration equation. Hence our next step is to see if we can reject the null hypothesis that there is no cointegration relation. The result of Johansen's (1988) test is sensitive to the length of lags included. Recall that this test is based on the VAR model. In order to find the optimum lag length, I follow several most commonly used criterion based on VAR model estimation[12]. The test results are shown in Table 9. The trace-stat is much larger than the 5% critical value for all the three groups, suggesting that we can reject the null hypothesis that there is no cointegration relation between these two series. Taking these two series as mutually cointegrated, by the Granger Representation Theorem (Engle and Granger (1987)), these cointegrated series can be formulated as a vector error correction (VEC) model of order 1. Specifically, consider a VAR model with p lags

$$\boldsymbol{y_t} = \boldsymbol{v} + \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{A}_2\boldsymbol{y}_{t-2} + \cdots + \boldsymbol{A}_p\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t \quad (2)$$

where $\boldsymbol{y_t}$ is an $N\,X\,1$ vector of variables, $\boldsymbol{v}$ is an $N\,X\,1$ vector of deterministic components, $\boldsymbol{A}_1$-$\boldsymbol{A}_p$ are $N\ X\ N$ matrices of coefficients, and $\boldsymbol{\varepsilon}_t$ is a $N\ X\ 1$ vector of residuals. We have two variables in $\boldsymbol{y}$, commonality and group liquidity. Therefore N=2

Then this VAR(p) model can be rewritten as the VECM:

$$\Delta\boldsymbol{y_t} = \boldsymbol{v} + \boldsymbol{\Pi}\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1}\boldsymbol{\Gamma_i}\Delta\boldsymbol{y_{t-i}} + \boldsymbol{\varepsilon}_t \quad (2.1)$$

where Δ is the differencing operator, $\boldsymbol{\Pi}$=$\sum_{j=1}^{p}\boldsymbol{A_j} - \boldsymbol{I_k}$ and $\boldsymbol{\Gamma_i}$= -$\sum_{j=i+1}^{p}\boldsymbol{A_j}$. When there is cointegration relation, the rank of the matrix $\boldsymbol{\Pi}$ has to be greater than 0 and smaller than N. In our case, the rank is 1. To facilitate analysis, let me write the model as

$$\Delta\boldsymbol{y_t} = \boldsymbol{v} + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1}\boldsymbol{\Gamma_i}\Delta\boldsymbol{y_{t-i}} + \boldsymbol{\varepsilon}_t \quad (2.2)$$

where $\boldsymbol{\alpha}$ is a $2\,X\,1$ vector, usually called the matrix of adjustment coefficients, $\boldsymbol{\beta}$ is a $2\,X\,1$ vector as well, usually called the cointegrating vector. Equation (2.2) is the baseline model and the coefficient matrices are to be estimated. However, we are facing an identification

[12] The most commonly used criterion to determine lag length include the Akaike information criterion (AIC), the Schwarz information criterion (SBIC) and the Hannan-Quinn information criterion (HQIC)

problem in that the estimation of parameters in α and β requires one identification restriction. I impose the most commonly used Johansen's (1995) normalization restriction and make the first parameter in β unity. Now we are ready to estimate the vector error correction model specified in (2.2).

Table 10 presents the estimation results for different size groups. The liquidity measure is scaled up by the order of 10^6 to make the results easier to read. Table 10 shows that the temporary change in the group liquidity and commonality is primarily explained by its lagged changes while the lagged changes in the other variable provide little explanatory power, which is consistent with the finding in Table 7. This result is robust for all three groups. Also note that the variables' own lags play a more important role in predicting future values for medium and large stocks. Moreover, on average the coefficients α and β estimated for these two variables are never significant simultaneously. For example, the estimated β in the commonality regression for small stocks is very significant (-5.06 with a t-stat of 5.57) but the estimated α in the liquidity regression is insignificant from zero (-0.01 with a t-stat of 0.08). This observation is valid for the other two groups, which implies deviation of one variable from equilibrium level does not cause the other variable to move to the new equilibrium level. Whereas, the opposite signs of α and β within each regression suggest that a variable is not likely to deviate too much from the equilibrium level. When that happens, it will move in the opposite direction in the next period.

The analysis above indicates that for a group of stocks, the time-series relation between its liquidity level and order flow commonality is not significant. There is little evidence for a Granger-causality relation between these two series. This implies that time variation of order flow commonality is more related to other group-level factors or some macro-level factors. In next section, I will show that commonality interacts more with aggregate market liquidity.

5.3 Commonality and Aggregate Market Liquidity

In this section, I examine the relation between commonality and aggregate market liquidity. Before proceeding, I need to make conjectures of the relation to form hypotheses.

Intuitively, order flow commonality among small stocks is positively related to the stock market liquidity. There are at least two reasons. First, when the market liquidity is high, investors are likely to be attracted to trade in the stock market. Therefore stock market, as a whole, will be a good place for informed investors to trade according to news and thus will have higher commonality in order flows. This can be called a “wealth effect”. Second, when the market liquidity is high, informed investors will be more willing to diversify their portfolios by investing in small and illiquid stocks and when market liquidity is scarce they will stay away from small stocks and trade large and liquid stocks. This is a “substitution effect”. Both of these two effects suggest a positive relation between commonality among small stocks and aggregate market liquidity. However, recall that we find that the order flow commonality of small stocks is not very volatile and does not seem to be as sensitive to underlying economic factors as that of medium and large stocks (Figure 4). The actual relation remains to be seen in later analysis.

The hypothetic relation between the order flow commonality of large stocks and market liquidity is not as clear. On the one hand, the wealth effect suggests there should be an increase (decrease) in commonality for the whole stock market if market liquidity is high (low), which suggests a positive relation. On the other hand, when market liquidity flourishes informed investors are more willing to diversify their portfolio and trade small stocks in response to economic news. For example, firms may choose to use large and liquid stocks to hedge when liquidity is low but are more willing to use a broader set of assets including small and illiquid stocks to increase return correlation with that of underlying assets when liquidity is high. The wealth effect implies a positive relation whereas the substitution effect implies a negative effect. Hence the net relation is not very clear. The question as to which effect is dominant needs to be answered by our time series analysis. To test these hypotheses, I use the same technique discussed in Section 5.2. I estimate a VAR model to test for Granger-Causality between commonality and aggregate market liquidity. Then I test for unit root of market liquidity, cointegration relations and choose an appropriate model to estimate regression coefficients and Granger-Causality.

The VAR estimation results reported in Table 11 suggest that commonality is more related to aggregate market liquidity than to group liquidity. First of all, all the regressions with commonality on the left hand side have higher R-squared compared with their counterparts in Table 7 (for example, 10.63% versus 6.29% for small stocks). Lagged commonality plays a significant role in explaining movements in the market liquidity for medium and large stocks. However, we observe both positive effects and negative effects. As discussed above, if we interpret the positive effect as wealth effect and the negative effect as substitution effect, we find that wealth effect leads substitution effect in time. In other words, if the stock market is going to experience a positive liquidity shock, informed investors will first be attracted to stock market, and they will further diversify their portfolio. Note that the negative substitution effect is larger in magnitude. Table 11 also suggests that the movements in market liquidity lead movements in commonality of small stocks, which implies that movements in commonality of large stocks can reflect new information faster than those for small stocks. Table 12 presents the estimates of the impulse response functions.

An untabulated augmented DF test yields a p-value of 0.256 (0.304 when a time trend included), which suggests that we cannot reject the hypothesis that there is a unit root in the market liquidity. Therefore I still employ Johansen's (1988) test to examine cointegration relations. The length of lags is selected following the same rule as above. Table 13 presents the test results, which indicate that we can reject the null hypothesis that there is no cointegration relation for all 3 size groups. Thus the estimation technique will be the same as that I use in last section.

Table 14 reports the results from VECM estimation. It can be seen that results are very different from those in Table 10. First note that unlike the group liquidity, the changes in market liquidity level cannot be predicted simply by looking at past changes in itself. The coefficients on its own lags are not significant for all of the 3 groups, which suggests that the substantial R-squared for medium and large groups (13.8% and 14% respectively, in contrast only 3.6% for small group where lagged change in commonality plays an insignificant role) is mainly attributed to lagged changes in order flow commonality. This finding confirms that the time series of commonality is closely related to market liquidity. Next, consistent with results in Table 10, I find lagged changes in order flow commonality have Granger-causality effect on the market liquidity change but not the other way round, primarily for medium and large stocks. The fact that the lagged changes in commonality lead the changes in market liquidity suggests that investors take actions before market liquidity changes and their trading has predictive power for the market liquidity change. Specifically, the positive coefficients on lagged change in commonality for medium and large stock groups indicate that when order flow commonality increases in these stocks, market liquidity level is going to decrease in one or two months. This is consistent with the "flight to liquidity" story argued above. Similar to Table 10, there is evidence for the wealth effect as well. For large stocks, the estimated adjustment coefficient α in the market liquidity regression is negative and significant (at 1% level) and the β in the commonality regression is positive and significant (at 10% level). These two coefficients, put together, suggest that given the market liquidity level, if the order flow commonality of large stocks is too high relative to the equilibrium level, then market liquidity is expected to increase in the next period to accord with the commonality level at a new equilibrium. However, the magnitude of the wealth effect is about -0.08 X 0.22=0.018, which is much less than the substitution effect coefficient 0.06. Thus we conclude that the substitution effect is dominant and a negative relation is revealed for both medium and large stocks.

In this section, I have investigated how the time series of order flow commonality for a group of stocks is related to the average stock liquidity in that group and the aggregate market liquidity. I find that commonality is more closely related to the market liquidity rather than to the group liquidity. Granger-causality of commonality on market liquidity is found primarily for medium and large stocks. Specifically, an increase in order flow commonality among medium and large stocks predicts a decrease in the market liquidity in one or two months, which reflects that informed investors prefer to trade large and liquid stocks when they expect market liquidity is going to decrease.

6. Commonality and the VIX

An interesting observation from above analysis is that order flow commonality shocks lead aggregate market liquidity shocks in time, which implies that examining investors' trading patterns can help researchers predict the changes in other financial variables. The time variation of order flow commonality is still an unexplored area in finance studies. If investors' trading activities contain information for future movements of other variables, obviously it is important to understand how to take advantage of the data on commonality movements. In this section, I examine a natural connection between commonality and the Chicago Board Options Exchange Market Volatility Index (VIX). Commonality represents investors' contemporaneous trading in different stocks in response to common economic information, and the VIX is the implied volatility of S&P 500 index options, which usually serves as a proxy for market uncertainty. However, the empirical relation between these two variables is not studied yet. A simple calculation of the contemporaneous correlation coefficient between the order flow commonality and the average VIX in a month shows that VIX is highly positively correlated with commonality among large stocks (coefficient=0.52). But no significant correlation coefficient is found for small and medium stocks (coefficient=0.01 and 0.09 respectively). This is not surprising since VIX is computed based on large stocks only. Also this suggests that higher market uncertainty is associated with more common movements in large stocks. To investigate the Granger-Causality relation, again I run a simple VAR model between commonality and the VIX. Table 15 presents the estimation results. I find that commonality of all 3 size groups has Granger-Causality effect on the market uncertainty measure VIX. Specifically, an increase in order flow commonality for medium and large stocks predicts an increase in market uncertainty in one month while an increase in commonality for small stocks predicts a decrease in market uncertainty in two months. This is consistent with the hypothesis that investors tend to trade large and liquid stocks when expected market uncertainty is low and trade small and illiquid stocks when expected market uncertainty is high. Impulse response functions are summarized in Table 16 and lead to similar conclusions.

Table 17 reports the multivariate VAR results including the return commonality as a control variable. Return commonality is measured based on intra-day stock returns using the same method as the order flows commonality measure. The inclusion of return commonality as does not change the explanatory power of the order flows commonality among large and liquid stocks.

7. Conclusion

This paper starts by asking the question how order flow commonality is related to stock liquidity and investigates the connection between these two variables in both panel and time series analysis. It is shown that cross-sectional sector-size groups with higher liquidity have significantly higher order flow commonality than groups with lower liquidity. The effect of liquidity on commonality is robust and provides more explanatory power than other documented factors such as stock indexing and institutional trading. Although the cross sectional relation between order flow commonality and group liquidity is strong, there is little evidence that a group's commonality is closely related to its own stock liquidity level on time dimension. Instead, the variation of commonality over time is more related to the aggregate market liquidity, primarily for medium and large stocks. In particular, an increase in the commonality of medium and large stocks predicts a decrease in market liquidity one or two months ahead. Then I further investigate the connection between commonality and market uncertainty, measured by the VIX. I find VIX is highly positively correlated with commonality among large stocks. An increase in order flow commonality for medium and large stocks predicts an increase in market uncertainty in one month while an increase in commonality for small stocks predicts a decrease in market uncertainty in two months. The results of my analysis suggest that investors prefer to trade large and liquid stocks when the expected market liquidity is low and the expected uncertainty is high.

More important, it shows that commonality in order flows contains important information about the market and is a useful tool for predicting future movements in other financial variables. To have better understanding of the stock market, further research on how to utilize the information in commonality is needed.

References

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

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5, 31–56.

Barberis, N., Shleifer, A., 2003. Style investing. *Journal of Financial Economics* 68, 161–199.

Barberis, N., Shleifer, A., Wurgler, J., 2005. Comovement. *Journal of Financial Economics* 75, 283–317.

Beber, A., Brandt, M., Kavajecz, K., 2011. What Does Equity Sector Orderflow Tell us About the Economy? *Review of Financial Studies* 24, 3688-3730

Beck, N. and Katz, J., 1995. What to Do (and Not to Do) With Time-Series Cross-Section Data. *American Political Science Review* 89:634-647.

Brandt, M. W., and K. A. Kavajecz. 2004. Price Discovery in the U.S. Treasury Market: The Impact of Order Flow and Liquidity on the Yield Curve. *Journal of Finance* 59, 2623–54.

Chan, L., Lakonishok, J., 1993. Institutional trades and intraday stock price behavior. *Journal of Financial Economics* 33, 173–200.

Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. *Journal of Financial Economics* 56, 3–28.

Chordia, T., Roll, R., and Subrahmanyam, A., 2005, Evidence on the Speed of Convergence to Market Efficiency, *Journal of Financial Economics* 76, 271-292.

Chordia, T., Subrahmanyam, A., 2004. Order imbalance and individual stock returns. *Journal of Financial Economics* 72, 485–518.

Corwin, S., Lipson, M., 2011. Order Characteristics and the Sources of Commonality in Prices and Liquidity, *Journal of Financial Markets* 14, 47-81

Cochrane, J., 1991. [Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots]: Comment. *NBER Macroeconomics Annual*, Vol. 6, (1991), pp. 201-210

Dickey, D. A. and Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427—431.

Domowitz, I., Hansch, O., Wang, X., 2005. Liquidity commonality and return co-movement. *Journal of Financial Markets* 8, 351–376.

DellaVigna, S., and Pollet, J., 2009. Investor inattention and Friday earnings announcements, *Journal of Finance* 64, 709–749

Easley, D.; M. O'Hara; and P. S. Srinivas. Option Volume and Stock Prices: Evidence on where Informed Traders Trade. *Journal of Finance* 53, 431–465.

Engle, R., and Granger, C., 1987. Co-integration and error-correction: Representation, estimation and testing, *Econometrica* 55, 251—276.

Goyenko, C., Holden, C., and Trzcinka, C. Do liquidity measures measure liquidity? *Journal of Financial Economics* 92, 153–181.

Granger, C., 1969. Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* 37, 424—438.

Granger, C., 1981. Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* 16, 121—130.

Greene, W., 1999. Econometric Analysis. Prentice Hall.

Harford, J., Kaul, A., 2005. Correlated order flow: Pervasiveness, sources, and pricing effects. *Journal of Financial and Quantitative Analysis* 40, 29–55.

Hasbrouck, J., Seppi, D. J., 2001. Common factors in prices, order flows, and liquidity. *Journal of Financial Economics* 59, 383–411.

Hirshleifer, D..and S. Lim, and S. Teoh. 2009. Driven to Distraction: Extraneous Events and Underreaction to Earnings News. *Journal of Finance* 5, 2289-2325.

Hughen, J. C., McDonald, C. G., 2006. Does order flow commonality extend across trade sizes and securities? *Financial Management* 35, 107–128.

Hvidkjaer, S., 2006. A trade-based analysis of momentum. *Review of Financial Studies* 19, 457–491.

Hvidkjaer, S., 2008. Small trades and the cross-section of stock returns. *Review of Financial Studies* 21, 1123-1151.

Johansen, S., 1988. Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231—254.

Johansen, S.: 1995, Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press, Oxford.

Kaniel, R., Saar, G., Titman, S., 2008. Individual investor trading and stock returns. *Journal of Finance* 63, 273–310.

Korajczyk, R. A., Sadka, R., 2008. Pricing the commonality across alternative measures of liquidity. *Journal of Financial Economics* 87, 45–72.

Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. *Journal of Finance* 46, 733–746.

McInish, T. H., Wood, R. A., Ord, J., 1985. An investigation of transactions data for NYSE stocks. *Journal of Finance* 40, 723–739.

Pastor, L., and Stambaugh, R., 2003. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy* 111:642–85.

Schwert, W., (1989). Test for Unit Roots: A Monte Carlo Investigation, *Journal of Business and Economic Statistics* 7, 147-159.

Table 1

PCA Estimation Results of Commonality in Stock Order Flows

	Eigen Values by GICS Sector Code					
	10			15		
Market Cap	1st	2nd	3rd	1st	2nd	3rd
Small	1.54	1.4	1.33	1.63	1.47	1.35
Medium	1.76	1.51	1.49	2.12	1.8	1.46
Large	2.06	1.58	1.49	1.72	1.51	1.47
	20			25		
	1st	2nd	3rd	1st	2nd	3rd
Small	1.5	1.42	1.4	1.45	1.43	1.4
Medium	2.04	1.43	1.35	1.9	1.56	1.41
Large	2.02	1.54	1.37	1.73	1.57	1.47
	30			35		
	1st	2nd	3rd	1st	2nd	3rd
Small	1.54	1.4	1.36	1.5	1.4	1.36
Medium	1.92	1.48	1.38	1.7	1.43	1.38
Large	1.62	1.49	1.44	1.77	1.6	1.46
	40			45		
	1st	2nd	3rd	1st	2nd	3rd
Small	1.62	1.49	1.41	1.49	1.44	1.32
Medium	2.25	1.42	1.37	1.81	1.56	1.45
Large	1.76	1.54	1.46	1.98	1.66	1.47

Note: This table presents the estimated eigenvalues of the first three components. The sample covers 8 GICS sectors. There are 30 randomly selected stocks in each sector-size group. The Principal Component Analysis is conducted for each group at the monthly frequency and eigenvalues are presented as time series average. The interpretation of GICS code is as follows. 10: energy, 15: materials, 20: industrials, 25: consumer discretionary, 30: consumer staples, 35: health care, 40: financials, 45: information technology.

Table 2

% of Variance Explained by the First 3 Principle Components

	GICS Sector Code							
	10	15	20	25	30	35	40	45
Market Cap	Percent of Total Variance Explained by First 3 Principal Components							
Small	14.45%	14.41%	14.52%	14.37%	14.50%	14.31%	14.43%	14.52%
Medium	14.58%	15.14%	14.57%	14.57%	14.51%	14.62%	14.78%	14.84%
Large	15.32%	15.23%	14.79%	15.08%	15.32%	15.05%	14.85%	16.29%
Average	14.78%	14.93%	14.63%	14.68%	14.77%	14.66%	14.69%	15.21%

Note: This table presents the % of total variance of standardized order flows explained by the first three principal components. The sample covers 8 GICS sectors. There are 30 randomly selected stocks in each sector-size group. The Principal Component Analysis is conducted for each group at monthly frequency and eigenvalues are presented as time series average. The interpretation of GICS code is as follows. 10: energy, 15: materials, 20: industrials, 25: consumer discretionary, 30: consumer staples, 35: health care, 40: financials, 45: information technology.

Table 3

Time Series Correlation Coefficients of Orderflow Commonality with Market Liquidity

	Market Liquidity
Sector	Correlation
Energy	-0.259***
Material	-0.153*
Industrial	-0.170**
Consumer Discretionary	-0.018
Consumer Staple	-0.006
Health Care	0.047
Financials	-0.135*
Info Tech	0.006

Note: This table reports the correlation coefficients between the two time series order flow commonality and the aggregate market liquidity for different sectors. Monthly commonality is measured by the first eigenvalue from PCA. * significant at 10% level, ** at 5% level and *** at 1% level.

Table 4

Estimation Results of Beck and Katz (1995) Regressions

	Beck and Katz Estimation				
	(1)	(2)	(3)	(4)	(5)
	COM	COM	COM	COM	COM
Groupliq	0.0268***	0.019***	0.010***	0.007***	0.011***
	(0.003)	(0.003)	(0.001)	(0.001)	(0.002)
Markliq		559.285	-61.693	1334.000	3584.020
		(1335.829)	(808.921)	(1789.790)	(2968.526)
Index		0.006***	0.001**	0.001*	0.001
		(0.001)	0.000	(0.001)	0.000
COM_{t-1}			0.691***	0.576***	0.508***
			(0.021)	(0.020)	(0.032)
Size				0.003*	0.004**
				(0.002)	(0.002)
# of segments				0.0253***	0.051***
				(0.006)	(0.009)
IO					-0.013
					(0.015)
Time and Group fixed effects	NO	NO	YES	YES	YES
constant	4.275***	4.283***	1.311***	2.528***	2.005***
	(0.010)	(0.016)	(0.090)	(0.152)	(0.136)
observations	3744	3744	3720	3720	1915
R-squared	19.16%	19.87%	47.34%	50.94%	56.89%

Note: This table reports the estimation results using Beck and Katz (1995) method. Dependent variable is the order flow commonality among stocks in each sector-size group. Sample period covers 1993-2005, with 24 groups in each year and 3744 observations in total. Sample size is reduced due to data availability on institutional ownership. Liquidity is measured using Amihud price impact and equally averaged across stocks in a group. Groupliq is the decile number of average stock liquidity in that group within my sample (higher number represents higher liquidity). Markliq is the aggregate market liquidity measure, which is equal-weighted average across all stocks in the market. Index is the number of stocks in S&P index for a group and COM_{t-1} is the lagged commonality. Size is the decile number of average market capitalization and # of segments is the average number of firm segments. IO is the average institutional ownership. *significant at 10% level, ** at 5% level and *** at 1% level.

Table 5

An Event Study: Non-Farm Payroll Employment

Size	Order Flow Commonality				
	(1)	(2)	(3)		
	Post-announcement Two Week Average	Long-run Mean	Pre-event 2 Week Average	Diff (1)-(2)	Diff (1)-(3)
Small	18.16	18.42	18.40	-0.26*	-0.24
Medium	18.80	18.53	18.29	0.26	0.51**
Large	20.21	19.74	19.67	0.47*	0.54***

Note: This table presents the stock order flow commonality (measured by the sum of first three eigenvalues) for different size for the 2 weeks after the non-farm payroll announcement and compare it with the long-run mean and the 2 weeks' prior to the event. * significant at 10% level, ** at 5% level and *** at 1% level

Table 6

Augmented Dickey-Fuller Test for Unit Root

Group	Z-stat	
	Group liquidity	Commonality
Small	-1.74	-2.35
	(0.41)	(0.16)
Medium	-2.42	-0.35
	(0.14)	(0.92)
Large	-1.47	-1.78
	(0.55)	(0.39)
# of Obs	142	142

Note: This table presents the Augmented Dickey-Fuller test for unit roots for the time series of group liquidity and order flow commonality of different size groups. P-value is reported in the brackets. H_0: A unit root is present. # of lags=13

Table 7

VAR Estimation of Relation between Group Liquidity and Commonality

	Small		Medium		Large	
	Groupliq	Com	Groupliq	Com	Groupliq	Com
$Groupliq_{t-1}$	0.361***	-0.074	0.422***	-0.953	0.478***	-1.348
	(0.082)	(0.170)	(0.071)	(1.479)	(0.081)	(5.274)
$Groupliq_{t-2}$	.416***	0.066	0.610***	0.090	0.143	1.515
	(0.081)	(0.167)	(0.065)	(1.370)	(0.089)	(5.822)
$Groupliq_{t-3}$	0.075	0.036	-0.284***	-0.477	0.040	-6.703
	(0.079)	(0.164)	(0.069)	(1.441)	(0.080)	(5.249)
Com_{t-1}	0.093**	0.077	-0.001	0.232***	0.002	0.386***
	(0.040)	(0.082)	(0.004)	(0.076)	(0.001)	(0.080)
Com_{t-2}	-0.010	0.162*	0.002	0.196**	0.000	0.267***
	(0.040)	(0.083)	(0.004)	(0.077)	(0.001)	(0.084)
$Comt_{-3}$	-0.032	0.125	-0.003	0.490***	-0.004	0.226***
	(0.040)	(0.082)	(0.004)	(0.076)	(0.001)	(0.081)
Constant	-0.388	5.809***	0.036	0.950*	-0.002	1.533**
	(0.555)	(1.150)	(0.026)	(0.539)	(0.010)	(0.658)
R-squared	69.84%	6.29%	68.65%	71.54%	37.62%	60.74%
Obs	153	153	153	153	153	153

Note: This table presents the estimation results of the vector autoregression model for 3 size groups. The two time series variables are group liquidity (illiquidity in fact) and order flow commonality. Groupliq is the group liquidity (Amihud price impact measure). Com is the commonality measure (the sum of first three eigenvalues in PCA). Standard errors are reported in the brackets. * significant at 10% level, ** at 5% level and *** at 1% level.

Table 8
Impulse Response Estimation between Group Liquidity and Commonality

	Small		Medium		Large	
	Groupliq	Com	Groupliq	Com	Groupliq	Com
$Groupliq_{t-1}$		-0.019		-0.438		-0.244
		(0.043)		(0.681)		0.953
$Groupliq_{t-2}$		0.008		-0.245		0.063
		(0.039)		(0.613)		(1.008)
$Groupliq_{t-3}$		0.003		-0.690		-1.211
		(0.027)		(0.516)		(0.943)
Com_{t-1}	0.726***		-0.013		0.032	
	(0.209)		(0.038)		(0.016)	
Com_{t-2}	0.498**		0.003		0.026	
	(0.253)		(0.044)		(0.016)	
$Comt_{-3}$	0.436		-0.033		0.011	
	(0.266)		(0.035)		(0.011)	
Obs	153	153	153	153	153	153

Note: This table presents the estimation results of impulse response funtion based on the VAR model in Table 7 for 3 size groups. The two time series variables are group liquidity (illiquidity in fact) and order flow commonality. Groupliq is the group liquidity (Amihud price impact measure). Com is the commonality measure (the sum of first three eigenvalues in PCA). Standard errors are reported in the brackets. * significant at 10% level, ** at 5% level and *** at 1% level.

Table 9

Johansen's (1988) Test for Cointegration of Commonality and Group Liquidity

	Ho: No cointegration relation		
Group	Trace Stat	Critical Value	# of lags
Small	53.94	15.41	2
Medium	27.53	15.41	3
Large	31.05	15.41	2

Note: This table presents the Johansen's (1988) Test for Cointegration relation between the time series of group liquidity and order flow commonality for different size groups. Trace-stat and critical value at 5% significance level is reported.

Table 10

VECM Estimation of Relation between Group Liquidity and Commonality

	Small		Medium		Large	
	*Δ*Groupliq	*Δ*Com	*Δ*Groupliq	*Δ*Com	*Δ*Groupliq	*Δ*Com
α	-0.01	0.13***	-0.25***	-1.24	-0.29***	-8.72*
	(0.013)	(0.026)	(0.050)	(1.060)	(0.055)	(5.214)
β	1.00	-5.06***	1.00	0.01	1.00	0.00
		(0.910)		(0.009)		(0.003)
$\Delta Groupliq_{t-1}$	-0.58***	-0.16	-0.32***	0.16	-0.30***	6.28
	(0.082)	(0.167)	(0.078)	(1.621)	(0.078)	(6.054)
$\Delta Groupliq_{t-2}$	-0.11	-0.12	0.28***	0.31	0.30***	7.07
	(0.082)	(0.084)	(0.070)	(1.475)	(0.070)	(5.432)
ΔCom_{t-1}	0.05	-0.28**	0.00	-0.73***	0.00	-0.57***
	(0.057)	(0.114)	(0.004)	(0.076)	(0.001)	(0.081)
ΔCom_{t-2}	0.04	-0.12	0.00	-0.51***	0.00	-0.27***
	(0.042)	(0.084)	(0.004)	(0.076)	(0.001)	(0.081)
Constant	-0.02	0.00	-0.15***	0.00	0.00	0.00
	(0.022)	(0.045)	(0.004)	(0.093)	(0.002)	(0.103)
R-squared	27.60%	46.70%	50.00%	40.90%	23.80%	26.60%
Obs	153	153	153	153	153	153

Note: This table presents the estimation results of the vector error correction model for 3 size groups. The two time series variables are group liquidity (illiquidity in fact) and order flow commonality. Groupliq is the group liquidity (Amihud price impact measure). Com is the commonality measure (the sum of first three eigenvalues in PCA). α is the adjustment coefficient and β is error correction coefficient. Johansen's normalization is imposed to ensure identification. Standard errors are reported in the brackets* significant at 10% level, ** at 5% level and *** at 1% level.

Table 11

VAR Estimation of Relation between Aggregate Market Liquidity and Commonality

	Small		Medium		Large	
	Markliq	Com	Markliq	Com	Markliq	Com
$Markliq_{t-1}$	0.798***	0.326*	0.804***	0.416	0.845***	0.418
	(0.082)	(0.181)	(0.079)	(0.319)	(0.080)	(0.400)
$Markliq_{t-2}$	0.038	-0.448**	0.066	-0.559	0.017	0.008
	(0.104)	(0.228)	(0.101)	(0.411)	(0.103)	(0.515)
$Markliq_{t-3}$	0.056	0.263	0.037	-0.035	0.044	-0.599
	(0.077)	(0.169)	(0.074)	(0.300)	(0.074)	(0.370)
Com_{t-1}	0.052	0.029	0.058***	0.223***	0.051***	0.369***
	(0.037)	(0.082)	(0.019)	(0.075)	(0.016)	(0.080)
Com_{t-2}	0.028	0.137*	-0.021	0.184**	-0.049***	0.249***
	(0.037)	(0.082)	(0.019)	(0.077)	(0.017)	(0.085)
Com_{t-3}	-0.034	0.092	-0.041**	0.517***	-0.007	0.223***
	(0.037)	(0.082)	(0.019)	(0.076)	(0.016)	(0.082)
Constant	-0.348	6.662***	0.101	0.981*	0.112	2.040***
	(0.554)	(1.218)	0.128	(0.518)	(0.140)	(0.703)
R-squared	85.21%	10.63%	85.97%	72.16%	86.15%	61.53%
Obs	153	153	153	153	153	153

Note: This table presents the estimation results of the vector autoregression model for 3 size groups. The two time series variables are the aggregate market liquidity (illiquidity in fact) and order flow commonality. markliq is the market liquidity (Amihud price impact measure). Com is the commonality measure (the sum of first three eigenvalues in PCA). Standard errors are reported in the brackets. * significant at 10% level, ** at 5% level and *** at 1% level.

Table 12
Impulse Response Estimation between Market Liquidity and Commonality

	Small		Medium		Large	
	Markliq	Com	Markliq	Com	Markliq	Com
Markliqt-1		0.076*		0.097		0.098
		(0.043)		(0.075)		(0.094)
Markliqt-2		-0.042		-0.031		0.120
		(0.043)		(0.075)		(0.097)
Markliqt-3		0.04**		-0.030		0.004
		(0.019)		(0.036)		(0.060)
Com_{t-1}	0.070***		0.082***		0.070***	
	(0.025)		(0.019)		(0.025)	
Com_{t-2}	0.078**		0.062**		0.024	
	(0.290)		(0.025)		(0.027)	
$Comt_{-3}$	0.055		0.026		0.018	
	(0.031)		(0.020)		(0.024)	
Obs	153	153	153	153	153	153

Note: This table presents the estimation results of impulse response funtion based on the VAR model in Table 10 for 3 size groups. The two time series variables are group liquidity (illiquidity in fact) and order flow commonality. Groupliq is the group liquidity (Amihud price impact measure). Com is the commonality measure (the sum of first three eigenvalues in PCA). Standard errors are reported in the brackets. * significant at 10% level, ** at 5% level and *** at 1% level.

Table 13

Johansen's (1988) Test for Cointegration of Commonality and Aggregate Market Liquidity

	Ho: No cointegration relation		
Group	Trace Stat	Critical Value	# of lags
Small	122.66	15.41	1
Medium	21.72	15.41	2
Large	24.94	15.41	2

Note: This table presents the Johansen's (1988) Test for a cointegration relation between the time series of aggregate market liquidity and order flow commonality for different size groups. Trace-stat and critical value at 5% significance level is reported.

Table 14

VECM Estimation of Relation between Aggregate Market Liquidity and Commonality

	Small		Medium		Large	
	*Δ*Markliq	*Δ*Com	*Δ*Markliq	*Δ*Com	*Δ*Markliq	*Δ*Com
α	-0.01	0.22***	-0.09***	-0.19	-0.08***	-0.29**
	(0.019)	(0.041)	(0.031)	(0.125)	(0.029)	(0.145)
β	1.00	-3.39***	1.00	0.01	1.00	0.22*
		(0.513)		(0.012)		(0.121)
$\Delta Markliq_{t-1}$	-0.12	0.17	-0.10	0.58*	-0.06	0.59
	(0.08)	(0.173)	(0.077)	(0.315)	(0.078)	(0.396)
$\Delta Markliq_{t-2}$	-.007	-0.27	-0.03	0.00	-0.04	0.54
	(0.081)	(0.173)	(0.075)	(0.306)	(0.075)	(0.381)
ΔCom_{t-1}	-0.34	-0.23*	0.06***	-0.74***	0.06***	-0.53***
	(0.391)	(0.117)	(0.019)	('0.075)	(0.016)	(0.081)
ΔCom_{t-2}	0.03	-0.09	0.04**	-0.53***	0.01	-0.25***
	(0.039)	(0.083)	(0.019)	(0.077)	(0.016)	(0.083)
Constant	-0.21	-0.001	-0.05**	0.00	-0.04*	0.01
	(0.207)	(0.044)	(0.022)	(0.091)	(0.020)	(0.103)
R-squared	3.6%	49.1%	13.8%	42.4%	14.0%	28.3%
Obs	153	153	153	153	153	153

Note: This table presents the estimation results of the vector error correction model for 3 size groups. The two time series variables are aggregate market liquidity (illiquidity in fact) and order flow commonality. Markliq is the aggregate market liquidity (Amihud price impact measure) COM is the commonality measure (the sum of first three eigenvalues in PCA). α is the adjustment coefficient and β is error correction coefficient. Johansen's normalization is imposed to ensure identification.* significant at 10% level, ** at 5% level and *** at 1% level.

Table 15

VAR Estimation of Relation between VIX and Commonality

	Small		Medium		Large	
	VIX	Com	VIX	Com	VIX	Com
VIX_{t-1}	1.011***	0.002	0.992***	0.003	0.959***	0.053*
	(0.086)	(0.009)	(0.082)	(0.022)	(0.092)	(0.031)
VIX_{t-2}	-0.188	0.002	-0.229**	0.009	-0.232*	-0.058
	(0.119)	(0.013)	(0.115)	(0.039)	(0.120)	(0.040)
VIX_{t-3}	0.072	-0.012	0.146*	-0.031	0.168*	-0.010
	(0.083)	(0.009)	(0.080)	(0.021)	(0.086)	(0.029)
Com_{t-1}	-0.362	0.035	0.527*	0.297***	0.479*	0.351***
	(0.796)	(0.086)	(0.293)	(0.078)	(0.280)	(0.094)
Com_{t-2}	-2.160***	-0.016	-0.278	0.109	-0.218	0.371***
	(0.792)	(0.085)	(0.303)	(0.081)	(0.266)	(0.090)
Com_{t-3}	-0.751	0.024	-0.363	0.501***	-0.113	0.089
	(0.771)	(0.083)	(0.293)	(0.078)	(0.278)	(0.094)
Constant	12.646***	3.270***	2.190**	0.768**	1.300	1.250***
	(4.167)	(0.450)	(1.063)	(0.284)	(1.031)	(0.347)
R-squared	82.60%	4.77%	82.04%	62.63%	81.90%	51.97%
Obs	153	153	153	153	153	153

Note: This table presents the estimation results of the vector autoregression model for 3 size groups. The two time series variables are the VIX and order flow commonality. VIX is the Chicago Board Options Exchange Market Volatility Index, averaged over trading days in that month. Com is the commonality measure (the first eigenvalue in PCA). Standard errors are reported in the brackets. * significant at 10% level, ** at 5% level and *** at 1% level.

Table 16
Impulse Response Estimation between VIX and Commonality

	Small		Medium		Large	
	VIX	Com	VIX	Com	VIX	Com
VIXt-1		0.005		0.097		0.127
		(0.023)		(0.075)		(0.074)
VIXt-2		0.009		-0.031		0.028
		(0.022)		(0.075)		(0.078)
VIXt-3		-0.023		-0.030		-0.007
		(0.105)		(0.036)		(0.056)
Com_{t-1}	0.843**		1.262***		1.388***	
	(0.305)		(0.305)		(0.209)	
Com_{t-2}	0.039		0.964***		1.777***	
	(0.351)		(0.322)		(0.309)	
$Comt_{-3}$	-0.297		0.551**		1.372***	
	(0.378)		(0.272)		(0.350)	
Obs	153	153	153	153	153	153

Note: This table presents the estimation results of impulse response funtion based on the VAR model in Table 10 for 3 size groups. The two time series variables are VIX and order flow commonality. Com is the commonality measure (the first eigenvalues in PCA). Standard errors are reported in the brackets. * significant at 10% level, ** at 5% level and *** at 1% level.

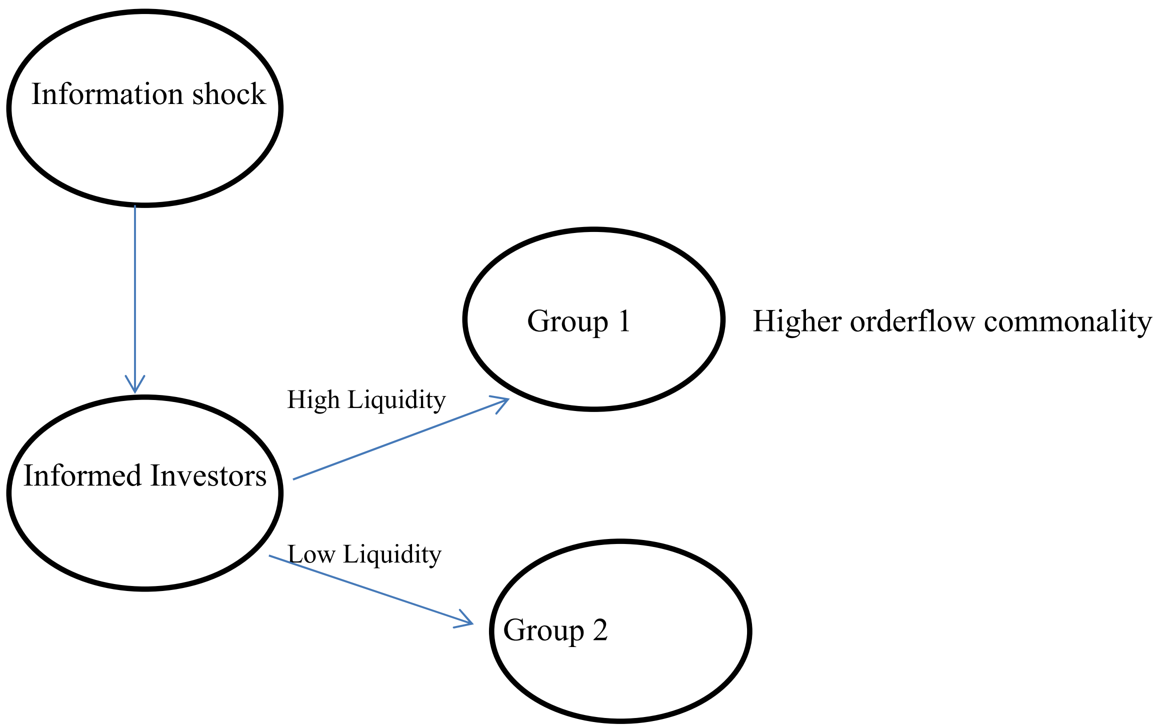
Table 17
VAR Estimation after Controlling for Return Commonality

	Markliq		VIX
$Markliq_{t-1}$	0.846***	VIX_{t-1}	0.947***
	(0.081)		(0.093)
$Markliq_{t-2}$	0.032	VIX_{t-2}	-0.228*
	(0.104)		(0.120)
$Markliq_{t-3}$	0.024	VIX_{t-3}	0.177**
	(0.075)		(0.087)
Com_{t-1}	0.052	Com_{t-1}	0.548*
	(0.016)		(0.286)
Com_{t-2}	-0.051	Com_{t-2}	-0.212
	(0.017)		(0.267)
Com_{t-3}	-0.027	Com_{t-3}	-0.130
	(0.017)		(0.279)
$Retcom_{t-1}$	0.000	$Retcom_{t-1}$	-0.001
	(0.000)		(0.005)
$Retcom_{t-2}$	0.000	$Retcom_{t-2}$	-0.001
	(0.000)		(0.005)
$Retcom_{t-3}$	0.000	$Retcom_{t-3}$	-0.005
	(0.000)		(0.005)
Constant	0.103		1.357
	(0.141)		(1.030)
R-squared	86.31%		82.04%
Obs	153		153

Note: This table presents the estimation results of the vector autoregression model for VIX and market liquidity after controlling for the return commonality for large and liquid stocks. markliq is the market liquidity (Amihud price impact measure). Com is the commonality measure. Retcom is the measure of return commonality. Standard errors are reported in the brackets. * significant at 10% level, ** at 5% level and *** at 1% level.

Figure 1

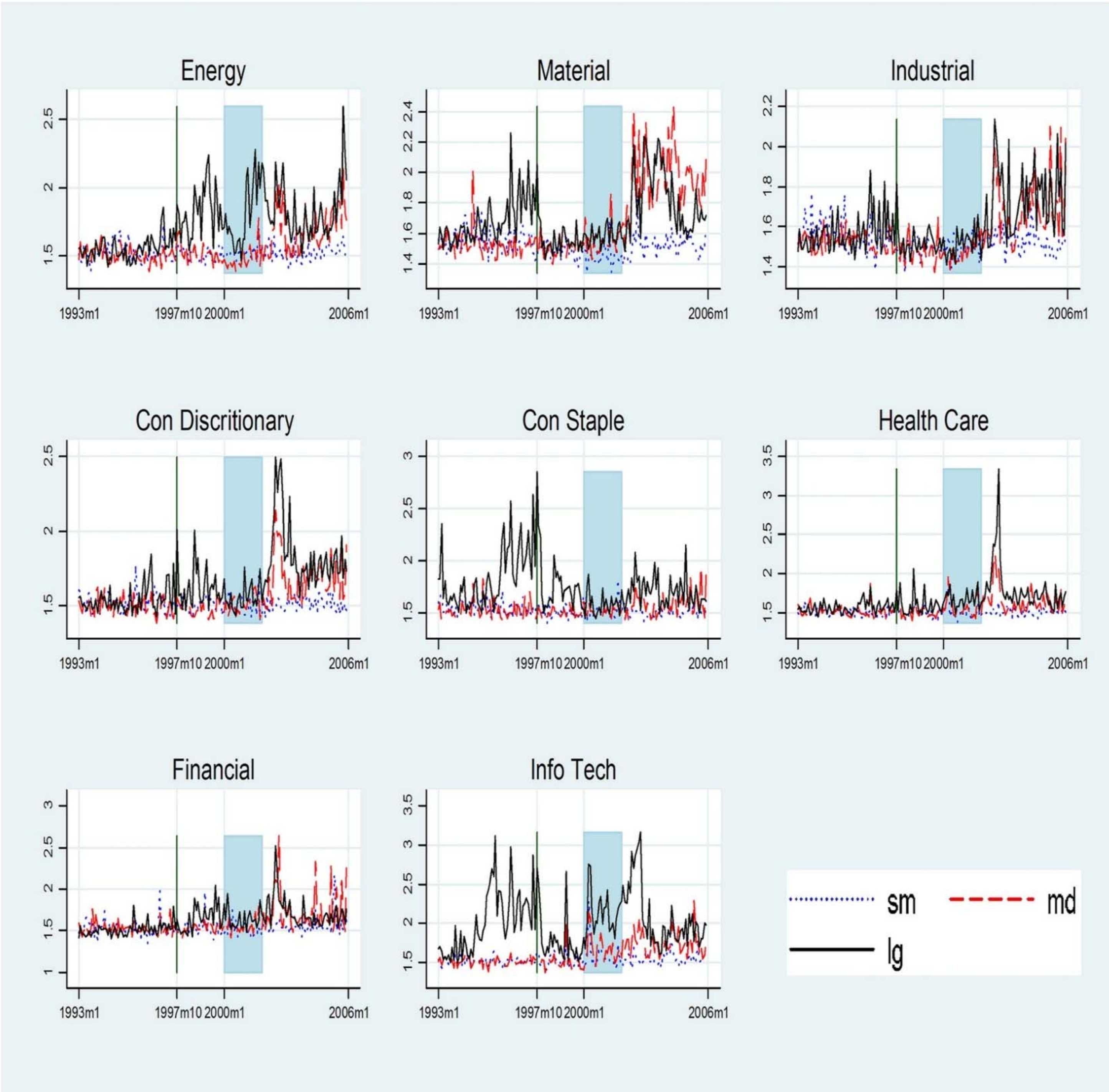
Order Flow Commonality and Stock Liquidity



Note: Figure 1 states that when investors receive an information shock that is common to some stocks or some investors, investors will choose stocks that have higher liquidity to reduce transaction cost. Therefore stronger commonality in order flows should be observed among more liquid stocks or in more liquid groups.

Figure 2

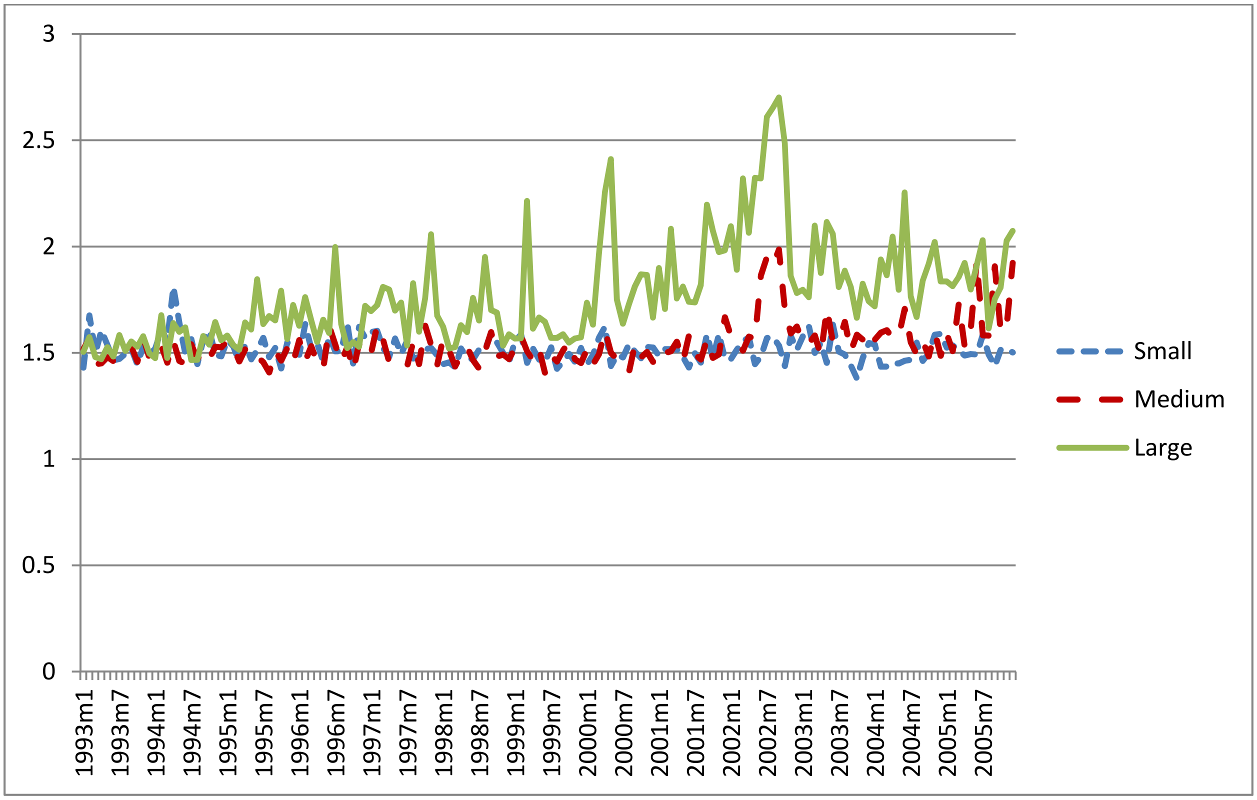
Time Series of Order flow commonality by Sector



Note: Figure 2 Plot the time variation of order flow commonality for stocks with different sizes in 8 different sectors

Figure 3

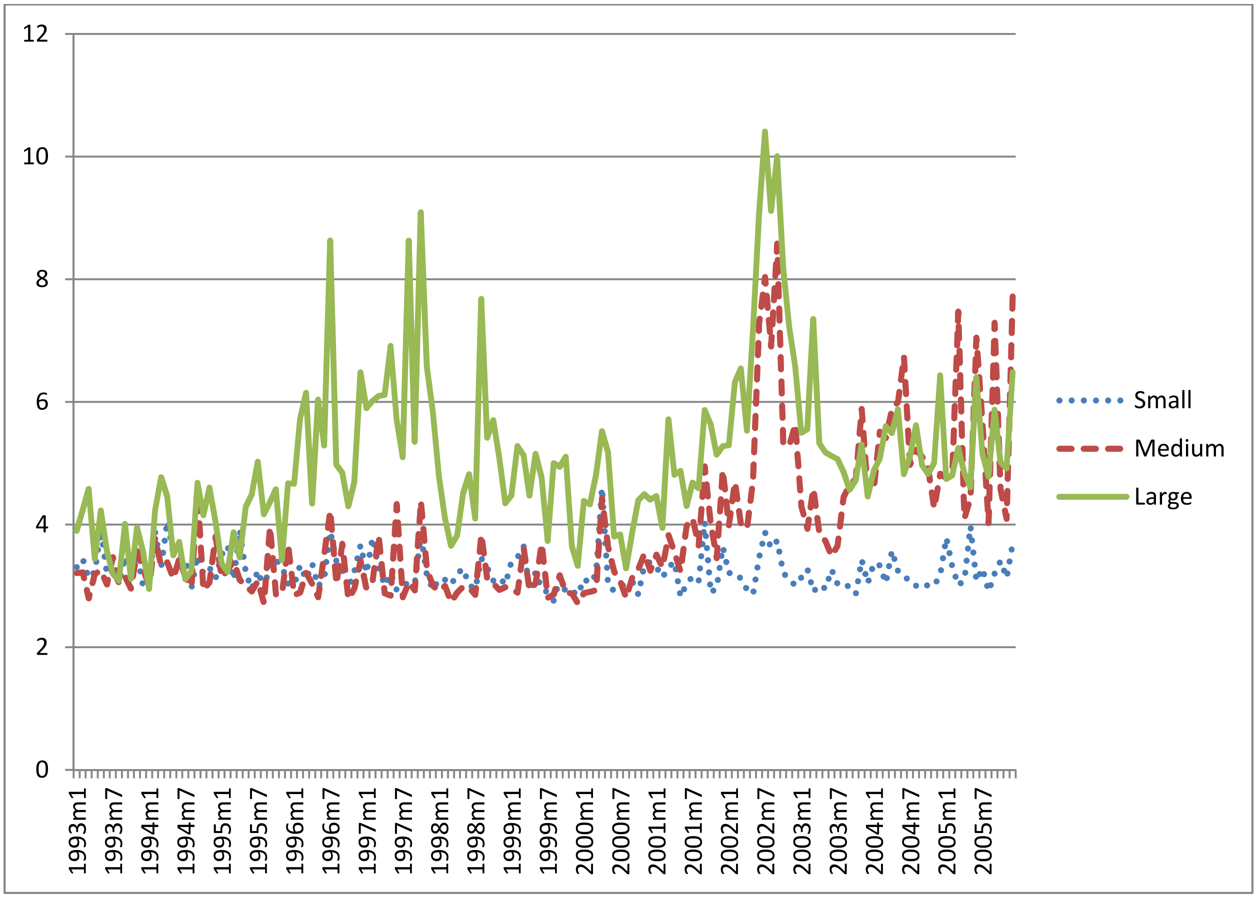
Order Flow Commonality for Stocks of Different Sizes: Standalone Firms Only



Note: Figure 3 presents the time variation of order flow commonality among small, medium and large stocks. Each size group consists of 80 standalone firms.

Figure 4

Order Flow Commonality for Stocks of Different Sizes



Note: Figure 4 presents the time variation of order flow commonality among small, medium and large stocks.