Trade Size and the Changing Nature of Price Formation

by

Ahmad Al-Haji

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Finance

Faculty of Business
University of Alberta

Abstract

Trading patterns in US financial markets have undergone significant changes in the past two decades. Using a 21-year (1993-2013) sample of intraday data, this thesis documents the ways in which the size distribution of trades—that is, the distribution of trades based on their dollar value—has changed over this period and examines changes in the price impact of trades and activities of informed traders.

Chapter 1 examines changes in trading activity and quantifies changes in the size distribution of trades between 1993 and 2013. On average, the daily trading volume per stock increased from about $2 m to $25 m, whereas the average dollar amount per trade decreased from over $40,000 to about $5,000 over the same period. In 1993, 75% of the trading volume came from large trades (in excess of $50,000 in value), but small trades (less than $5,000 in value) accounted for more than 40% of the volume in 2013.

The findings reported in Chapter 1 suggest the need for a study, presented in Chapter 2, which focuses on price formation over the sample period, contrasting the permanent and transitory price effects of trades conditional on their sizes. Changes in the price impact of trades are negatively related to trade size, with small trades exerting the largest price impact in recent years. Earlier studies such as that of Barclay and Warner (1993) showed that most "stealth" trading, i.e., strategic information-based trading, occurred in medium-sized trades. My results are consistent with those studies only in the early years of my sample period; they suggest that stealth trading now occurs in small trades. Further, the positive "price-quantity" relation predicted in Easley and O'Hara (1987) has seemingly vanished or even reversed in recent data.

The close association between the shift in trade size distribution and transition of permanent price impact, as demonstrated by results presented in Chapters 1 and 2, indicates that informed traders are directly involved in those change patterns. Chapter 3 analyzes in more detail the behavior of informed traders during my sample period, assessing their role in driving the findings reported in Chapters 1 and 2. I expect to find that the increase in small trading volume is associated with a decrease in medium trading volume in particular, since studies such as that of Barclay and Warner show that stealth traders tend to concentrate on medium-sized trades. The results of my test point in the direction of this conjecture. I also test whether a temporary increase in information-based trading shifts the distribution of trades toward smaller transactions. I classify stocks according to their probability of information-based trading (PIN) values during each quarter, and I find that stocks with high PIN values tend more often to be traded in small sizes. The findings reported in Chapter 3 suggest that informed traders are actively involved in the migration of trade volume toward smaller trade sizes.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Professor Aditya Kaul, for his continuous support throughout my years of study in the PhD Program. Our numerous long meetings and extended phone calls have always represented for me a source of motivation and renewed energy. These few words are inadequate to express my appreciation for Professor Kaul's care, generous support, guidance, and understanding. I am also grateful for the substantial support of the members of my thesis committee: Professor Valentina Galvani, Professor Vikas Mehrotra, Professor Akiko Watanabe, and Professor Masahiro Watanabe. I thank Professor Masahiro Watanabe for sharing various data sets and computing resources.

I am appreciative of help I received from many other faculty members, administrative staff, and friends. In particular, I thank Professor Lukas Roth for his role in my job market preparation. I also thank Jeanette Gosine and Debbie Giesbrecht from the Business PhD office, and my friends Pari Veeren and Xiaowen Zhang for their help and encouragement. Financial support arranged though the PhD office for the years 2012-2016 is acknowledged.

I owe my family many thanks for their emotional and financial support: my parents, Abderramàn El-Hajji and Manal Abdulkarim; my brother-in-law, Ali Alkhaja; my sisters, Ibtihal and Maab; and my brothers, Bilal and Aimen. I have a special message for my parents. Mom and Dad: You are my true inspiration; I am greatly indebted to you for all you have done for me… I love you!

بابا وماما: صنعتم الكثير من أجلي، شكراً لكما، أحبكما!

Table of Contents

List of Figures

List of Tables

Chapter 1 - An Overview of Recent Changes in Trading Activity

Introduction

The past two decades have seen huge changes in the functioning of financial markets and in trading patterns. The adoption of new technologies, changes in regulations, and the emergence of high frequency and algorithmic trading are associated with an explosion in trading volume and a reduction in trade size. Recent work in microstructure has examined aspects of these changes. For instance, Chordia et al. (2011) examined trends in market efficiency between 1993 and 2008; Goldstein and Kavajecz (2000), among others, studied the effects of changes in tick size; and Hendershott et al. (2011) investigated the effects of algorithmic trading on quote efficiency. Chapter 1 of this thesis contributes to this literature by quantifying the changes in the size distribution of trades using an extended sample period (1993-2013). The second and third chapters extend this contribution by examining changes in the price impact of trades and activities of informed traders.

Trade size is used in the microstructure literature as a proxy for trade type, which may be institutional or retail (e.g., Barber et al., 2009). Thus, my first question regarding changes in the size distribution of trades sheds light on changes in the market participation pool. In each trade size category I examine trading volume and information-based trading. I define order flow as the difference between buyer initiated- and seller-initiated trades. I measure trading activity by order imbalance because this measure conveys value-relevant information to markets (Chordia et al., 2011).

I aggregate imbalance at the daily level for the period 1993-2013. I start by documenting the explosion in trading volume in recent years. Accompanying this growth in trading activity is a sharp leftward shift in the distribution of trade sizes. For instance, comparing the two extreme years in my sample, I find that trading volume increased more than 10-fold during the sample period and that the average trade size dropped significantly, from $48,000 to $6,000. I show the shift in the distribution of trades over five size bins, following the thresholds used by Barber et al. (2009) to define bins. The proportion of small trading volume increased from 2% in the early years to 42% in recent years, whereas for large trading volume the proportion dropped from 75% to 18% over the same period. I find that most of the shift took place between 1998 and 2003. Moreover, prior to the late 1990s, the change in trading activity was very minimal. For instance, I find that average trade size in 1983 was about $53,000; this amount is just slightly higher than the average trade size in the early years of my sample.

This evidence complements that offered by Chordia et al. (2011) by showing exactly which trade sizes contribute to the lower average trade size. It also sets the stage for further tests examining the implications of this change for the behavior of informed traders and price impact of trades. The analysis in this study depends heavily on the accuracy of trade signing algorithms—the classification of trades as either buyer- or seller-initiated. In signing trades, I rely on a new approach recommended by Holden and Jacobsen (2014) for dealing with many issues associated with trade data, especially those emerging in more recent years. In addition, various signing methods have been proposed to sign trades. I consider three methods but I present results based on one of them—Lee and Ready's (1991) approach. A comparison of the performance of the three signing methods I consider is presented in Appendix 1.A.

The remainder of this chapter is organized as follows. Section 1.1 presents data sources. In Section 1.2, I explain the steps taken to process the data and design order flow measures. Section 1.3 presents results regarding the change in trading activity over the sample period, and Section

1.4 concludes the chapter. Appendix 1.A, which concerns the performance of signing methods, appears at the end of the chapter.

1.1 Data Sources

My data sets include common stocks listed on the NYSE, NASDAQ, and AMEX[1], and can be matched in both the Center for Research in Security Prices (CRSP) and the Trades and Quotes (TAQ) databases. I clean my CRSP data set as follows. Stocks with non-continuous observations or with fewer than 50 observations in any sample are removed from my data set for that sample. I also exclude observations pertaining to non-ordinary categories of stocks since the trading characteristics of these stocks are different than those of ordinary ones[2].

I take the following steps in cleaning the TAQ data set (Holden and Jacobsen, 2014). Trades and quotes outside hours are excluded. I disregard trades if the correction indicator shows that they have been revised; that is, I retain trades in the sample if the correction indicator equals "00". I also drop trades with non-positive or missing prices. I exclude quotes with abnormal modes, non-positive or missing bid or ask prices, negative spreads on the same exchange, spreads larger than $5, or non-positive or missing depth[3].

Although TAQ returns are expected to be identical to those appearing in CRSP, some differences can be noticed between the two return series. I try to limit any effect of errors in the TAQ files by applying a simple filter rule that deletes any transaction with an absolute return greater than 20%. I use the correlation between TAQ and CRSP returns as an indication of the

[1] My data are aggregated across NYSE, NASDAQ, and AMEX. These markets have different designs and protocols (Chordia et al., 2011). In an unreported analysis, I repeat most of my tests for NYSE and NASDAQ separately. Key findings are unaffected.

[2] Non-ordinary categories are as follows: Certificates, ADRs, shares of beneficial interest, units, companies incorporated outside the U.S., Americus Trust components, closed-end funds, preferred stocks, and REITs.

[3] More information about TAQ data fields can be found at: http://www.nyxdata.com/Data-Products/TAQ-Web.

efficacy of my TAQ filter. The overall cross-sectional average correlation between the two return series is 0.926[4]. The divergence between TAQ and CRSP returns can be attributed in part to differences in the closing prices[5], inaccurate TAQ-CRSP matching, or outright reporting mistakes in the TAQ. Nevertheless, the correlation reported above indicates that the two return series are generally very closely related. When I look at average correlations by year, I find that mean (median) annual correlations range from 0.87 (0.9) in 1993 and increase almost monotonically to 0.97 (0.98) in 2013. The fact that mean and median correlations are significantly higher in recent years indicates that TAQ data suffer from fewer inaccuracies in those years.

The data set covers a period of 21 years, which I split into two samples: An *early* sample for the period 1993-2001 and a *recent* sample for the period 2002-2013. The data set includes a wide cross-section of stocks and covers a lengthy period that has witnessed many changes in financial market design and trade regulations. The number of stocks in the final data set ranges from 4,217 in 2007 to 6,924 in 1997.

[4] I applied more sophisticated filtering rules on TAQ observations. The resulting returns are less correlated with CRSP returns. For instance, in an alternative filtering rule I discard any transaction with return (R_t) if it meets the following three conditions: (i) $|R_t| > x$, where x = 0.05, 0.10, … , 0.30; (ii) $|R_t * R_{t+1} -1| < y$, where y = 0.01 or 0.005; and (iii) $R_t * R_{t+1} < 0$. This three-condition rule attempts to detect transactions with substantial returns that immediately reverse. Since this situation is highly unlikely, such transactions are potentially erroneous and are removed by this filter. The resulting correlations between TAQ and CRSP returns are substantially lower than 0.926 for all combinations of the x and y values. In the light of this result, I decided to follow my simple and less conservative rule, which is condition (i) only with x = 0.2.

[5] In calculating the TAQ return, I assume that the closing price is the quote midpoint prevailing at 16:00. Roll (1984) and Kaul and Nimalendran (1990) show that if returns are calculated from transaction prices, they are biased by the bid-ask bounce.

1.2 Data Processing and Measures

I measure trading activity on the basis of dollar volume imbalance aggregated at the daily level and examine its impact on CRSP's daily return-excluding-dividends (*RETX*). Studies relating returns to trading activity have differed in the aggregation level they employed; these have ranged from intra-daily to annually. For instance, Harford and Kaul (2005) measure order imbalance and returns for 15-minute periods, whereas Barber et al. (2009) use weekly and even annual frequencies. I use *daily* data as a compromise that balances data manageability and endogeneity. I assume that prices respond to trading activity. However, as the measurement horizon is lengthened, the feedback channel from returns to trading activity becomes more of a concern (Harford and Kaul, 2005). In order to minimize this feedback effect, I do not aggregate my measures beyond the daily level, but I cannot rule it out completely. Therefore, my findings are subject to this caveat. In addition, the choice to analyze data at the daily level strikes a balance between the sharpness of identification of price impact and the accuracy of trade classification.

Examining imbalances at the intradaily level is expected to better capture the effect of order imbalance. On the other hand, the accuracy of trade classification into buyer- and seller-initiated trades improves with the level of aggregation. Chakrabarty et al. (2012) show that misclassification of trades is substantial at the trade level. They find that about 31% of trades are misclassified, but those misclassifications cancel out almost completely at the daily level. Since the credibility of my findings depends largely on the accuracy of trade classification, I decided to conduct my analysis at the daily frequency[6]. However, any remaining misclassification may

[6] Heston et al. (2010) show that return and order imbalance patterns are different depending on the frequency of data. For robustness, I perform most of the tests in this study at intradaily (30-minute) and monthly levels. Results from these analyses (unreported) are consistent with my main findings in this paper.

work only against my findings. Another important reason for using daily data is that the noise and reversal components of returns are not captured if measures are aggregated beyond the daily level (Chordia et al., 2002). In my analysis below, I regress returns on both contemporaneous and lagged imbalances. The latter is associated with a negative coefficient, which I interpret as reversal due to temporary price pressure or overreaction by uninformed traders the previous trading day. This reversal is very informative to my study because it allows me to isolate the permanent component of return related to information. Therefore, I examine how this reversal behaves throughout my analysis.

The trade and quote data from recent years present a challenge to microstructure researchers—one that extends beyond the requirement for extensive computing resources. During this period, multiple quotes and trades frequently took place within the span of each second. The presence of multiple data points within a second is a serious concern, especially for users of the whole-second TAQ database, which is typically used in academic research. A common procedure is to match all trades within a second to the *last* quote in the matching second. This approach clearly leads to inaccurate inferences about spreads, and more importantly for my case, about the direction of the trade (i.e. buyer- or seller-initiated). Other issues arising from the use of the whole-second TAQ include the presence of withdrawn quotes and quote cancellation without updating the database.

The best way to deal with this issue is to use the more expensive millisecond TAQ, which does not suffer from any of the above problems. However, for financially constrained researchers the less expensive way to deal with those issues is to apply corrective measures such as those proposed by Holden and Jacobsen (2014). Therefore, I follow their procedure in constructing the

National Best Bid and Offer (NBBO) quotes and signing trades[7]. The procedure can be summarized as follows. The authors adjust for withdrawn quotes that the whole-second TAQ records either as "zero" or as "missing" by excluding these quotes rather than replacing them with previous quotes on the same exchange, thereby avoiding the use of stale quotes. To overcome the absence of the millisecond timestamp in the whole-second TAQ, the order of trades and quotes within each second is used to approximate a millisecond timestamp. A trade at an approximate millisecond is matched to the NBBO quote prevailing in the prior approximate millisecond. In addition, the procedure attempts to infer cancelled quotes from negative or zero spreads on the NBBO and exchange levels, in which case associated quotes are excluded, as mentioned above.

I apply Holden and Jacobsen's (2014) procedure to three different trade signing algorithms, but my measures and analysis below are based on Lee and Ready's (1991) method[8]. Once NBBO is constructed, Lee and Ready's (LR, 1991) method of trade signing is followed, using a 0-second lag for trade data as recommended by Bessembinder (2003). The LR method classifies a trade as buyer- (seller-) initiated if the trade price is above (below) the prevailing midpoint of bid and offer quotes. If a trade occurs at midpoint, the trade price is compared to the prices of up

[7] I thank Craig Holden for making an SAS code of the recommended solution available on his website: http://www.kelley.iu.edu/cholden

[8] For robustness, I repeat my analysis using two alternative trade signing algorithms: those of Ellis et al. (2000) and Chakrabarty et al. (2007). In addition, I repeat my tests based on Lee and Ready's (1991) approach without applying Holden and Jacobsen (2014) adjustments, according to the procedure presented in the Wharton Research Data Services (WRDS) website, available at:
https://wrds-web.wharton.upenn.edu/wrds/research/applications/intraday/index.cfm
(a WRDS account is required to access the content of this webpage). I find that none of these methods has a significant impact on my results, and my findings remain unchanged. A discussion and comparison of the different algorithms that I use are presented in Appendix 1.A.

to two previous trades; if the price is above (below) the previous price, then the trade is considered to be a buyer- (seller-) initiated trade.

I use order imbalance to measure trading activity because a significant imbalance in either direction would entail actions by market makers to control inventory. These position adjustments by market makers are usually associated with significant changes in prices. Another advantage for imbalance is that it indicates investor interest in a stock, signaling potential information-based trading (Chordia and Subrahmanyam, 2004).

I refer to the number of shares traded, dollar volume, and number of trades for stock *i* at day *t* as VOL_{it}, $DVOL_{it}$, and N_{it}, respectively. I refer to dollar buy and sell volumes as $DVOLBUY_{it}$ and $DVOLSELL_{it}$, respectively. The numbers of buy- and sell-initiated trades are referred to as $NBUY_{it}$ and $NSELL_{it}$, respectively. I calculate order imbalance based on dollar volume and number of trades and refer to them as $DVOLIMB_{it}$ and $NIMB_{it}$, respectively; these variables are calculated as follows: $\frac{DVOLBUY_{it} - DVOLSELL_{it}}{DVOLBUY_{it} + DVOLSELL_{it}}$ and $\frac{NBUY_{it} - NSELL_{it}}{NBUY_{it} + NSELL_{it}}$. Note that the denominator is the volume that can be signed, given that a significant proportion of the volume is unsigned by the LR algorithm (Appendix 1.A).

In addition, I follow the definition provided by Barber et al. (2009) for trade size in classifying trades into trade size bins. Specifically, I use the following rules in classifying trades into bins: Bin 1 if $DVOL_{it} \leq \$5,000$ (small trades), Bin 2 if $\$5,000 < DVOL_{it} \leq \$10,000$, Bin 3 if $\$10,000 < DVOL_{it} \leq \$20,000$, Bin 4 if $\$20,000 < DVOL_{it} \leq \$50,000$, and Bin 5 if $\$50,000 < DVOL_{it}$ (large trades)[9]. I adjust these threshold values in each month by the ratio of the consumer

[9] TAQ rarely records odd-lot trades—trades with fewer than 100 shares (O'Hara et al., 2014). Therefore, some stocks with high prices tend to have few or no trades in small bins. The systematic absence of imbalance in certain bins for certain stocks might bias my findings. For robustness, I conduct most of the tests in this analysis on a subset

price index (*CPIAUCSL*) in that month to the average *CPIAUCSL* over the three months between November 1990 and January 1991[10]. Previous studies have classified trades into size bins on the basis of either the number of shares traded or the dollar volume of trades. I choose the latter approach because the dollar value incorporates the effects due to the level of stock price. A trade of 100 shares of a highly-priced stock is large but might be considered otherwise if a low-priced stock is involved. In the past, moreover, studies have classified trades into various numbers of size bins: two, three, or five[11]. I have opted to use five bins in order to obtain a clearer picture. My results show that a smooth pattern is discernable across these five bins.

Dollar volume and number of trades of stock i at day t in each bin j are referred to as $DVOLj_{it}$ and Nj_{it}, respectively, where $j = 1, \ldots, 5$. I aggregate signed dollar volume and add up the number of trades within each bin; then I calculate imbalances in each of the five trade size bins individually. I refer to buy and sell dollar volume (number of buy and sell trades) of stock i at day t in bin j as $DVOLBUYj_{it}$ and $DVOLSELLj_{it}$ ($NBUYj_{it}$ and $NSELLj_{it}$), respectively. Next, I calculate imbalances based on dollar volume ($DVOLIMBj_{it}$) and number of trades ($NIMBj_{it}$) in each bin j, respectively, as follows: $\frac{DVOLBUYj_{it} - DVOLSELLj_{it}}{DVOLBUYj_{it} + DVOLSELLj_{it}}$ and $\frac{NBUYj_{it} - NSELLj_{it}}{NBUYj_{it} + NSELLj_{it}}$.

of stocks that have trades in all trade size bins in each month. Results (unreported) show that my findings remain unchanged.

[10] Consumer price index data are obtained from the webpage of the Federal Reserve Bank of St. Louis. This is available at: http://research.stlouisfed.org/.

[11] For instance, Barclay and Warner (1993) classify trades based on the number of shares traded into three bins: small, medium, and large, for those involving less than 500 shares, between 500 and 9,900, and more than 10,000 shares, respectively. Easley et al. (1997) classify trades into two size bins based on the number of shares traded; they define small trades as those involving fewer than 1000 shares, whereas large trades are defined as those involving 1000 or more shares. Bessembinder and Kaufman (1997) classify trades into three size bins based on the dollar volume; bins are small if the dollar volume is less than $10,000, medium if the dollar volume is between $10,000 and $199,000, and large if the dollar volume is larger than $200,000. Finally, Chan and Fong (2000) classify trades into five size bins based on the number of shares traded with the following ranges: 1-500 shares, 501-1,000 shares, 1,001-5,000 shares, 5,001-9,999 shares, and more than 10,000 shares.

1.3 An Overview of Changes in Trading Activity

Table 1.1 presents variable means. Means are calculated first for each stock, and cross-sectional means are presented in the table. I present means for early (1993-2001) and recent (2002-2013) samples as well as for five sub-periods (1993-1996 [first], 1997-2000, 2001-2004, 2005-2008, and 2009-2013 [last]). Most of the subsequent analysis will make comparisons between early and recent samples only.

The remarkable increase in trading activity over time is obvious. Daily average dollar trading volume per stock has increased steadily more than 10-fold, from about $2 million in the first sub-period to approximately $25 million in the last sub-period, without adjusting for inflation. The rate of increase in number of trades is more pronounced: from 50 trades in the first sub-period to more than 4000 trades in the last sub-period (an 80-fold increase between the two sub-periods). The rate of increase in the number of trades is about eight times that of the dollar volume, as evidenced in the significant decrease in the average trade size from approximately $48,000 in the first sub-period to $6000 in the last sub-period.

Table 1.1 also illustrates the breakdown of average trade measures by trade size bins. Examining trading activity on this basis shows that trades of all sizes contributed to the overall increase in trading activity, but that most of the increase was driven by small trades (Bin 1). In terms of volume, small trades had the lowest share of the overall volume and large trades had the largest share of volume in the early years. The average daily volumes of Bin 1 and Bin 5 trades, respectively, were $46,624 and $1.792 million in the first sub-period. The higher rate of increase for small bins brought the volumes of small and large bins near to parity in the 2005-2008 sub-period. In the 2009-2013 sub-period, small bin volume reached $10.618 million, while that of large bins dropped to $4.668 million. In terms of the number of trades, the difference between small and large bins, as expected, was even more remarkable. The average number of trades per day in Bin 1 increased from 15 in the first subsample to 3,454 in the last subsample. Those numbers for Bin 5 are 8 and 17, respectively.

The fact that the rate of increase in small volume is higher than that for larger bins translates into a shift in the distribution of trades across bins. Specifically, the proportion of small trades has increased, whereas that of large trades has decreased, over the sample period. Quarterly average trading volume by bin is presented graphically in Figure 1.1. The two vertical lines refer to the period from the beginning of the fourth quarter of 2000 to the end of the second quarter of 2003. I examine changes carefully during this period since many regulatory and technological changes came into effect during these years. Those events include the reduction of the minimum tick size to a penny (decimalization) that came into full effect by January 29, 2001 for NYSE stocks and by April 9, 2001 for NASDAQ stocks. Prior to that, the NYSE implemented the Direct+ system on October 21, 2000, allowing for the automatic execution of trades involving up to 1,099 shares. Finally, this period witnessed the implementation of the Autoquote system, which was phased in gradually between January 29, 2003 and May 27, 2003. Hendershott et al. (2011) show that algorithmic trading has flourished subsequent to quote automation in 2003[12]. Subsequent graphs will highlight the same period.

To measure the extent of shift, Table 1.1 reports equally- and volume-weighted average proportions of trading volume in each bin, and Figure 1.2 plots those proportions on a quarterly basis. Figures show that the share of small trades gradually increased over the sample period,

[12] Hendershott et al. (2011) report that high-frequency trading (HFT), which is a subset of algorithmic trading, constituted 73% of trading volume in the United States. A more recent Bloomberg report shows that the HFT market share has gradually fallen since then to 50% in 2012, due to increased competition and tightening regulations. This Bloomberg report is available at: http://www.bloomberg.com/bw/articles/2013-06-06/how-the-robots-lost-high-frequency-tradings-rise-and-fall.

while that of large trades decreased. However, this pattern was somewhat reversed after 2010. During the highlighted period there was a major shift in the size distribution of trades. While Figure 1.1 shows that there was an increasing trend in trading volume in general, the highlighted period in particular did not witness significant changes in trading volume. The contrast between Figures 1.1 and 1.2 during this period shows that those changes in trading outcomes probably occurred in the absence of economic events.

The shift is obvious in both equally- and volume-weighted plots. However, the difference between the two is worth noting. In the case of volume-weighted average proportions of trading volume, large trades dominate heavily in the early years by capturing more than 60% of trading volume, while the share of small trades is the lowest, with less than 5%. With equally-weighted average proportions, large trades still have the highest share, but are closely followed by small trades. Shares for both bins are slightly above 25%. This difference shows that heavily-traded stocks tend to be traded in large sizes. Part of the reason is the higher liquidity levels that those stocks possess. Higher depths can absorb larger trades without a significant price impact. It is also possible that those heavily-traded stocks have less information asymmetry, and that there is therefore less need for strategic information-based trading in non-large sizes.

Average imbalances (*DVOLIMB* and *NIMB*) are also reported in Table 1.1, both in the aggregate and by bin. While several studies have shown that aggregate imbalances are positive on average (that is, there are more buyer-initiated than seller-initiated market orders), my results, shown in Table 1.1, demonstrate the opposite. This difference may be a function, in part, of the time period and particular markets on which I have focused. For instance, Chordia et al. (2002), who found that imbalances are positive on average, considered S&P500 stocks for the period 1988-1998, whereas my sample includes all NYSE, NASDAQ, and AMEX stocks for the period

1993-2013. Indeed, I find that my imbalance measure is generally positive when I restrict the analysis to S&P500 stocks, but that the magnitude is slightly smaller than that reported by Chordia et al. (2002).

The remaining discrepancy between my results and those of Chordia et al. (2002) and other, similar studies may be attributed to differences in the construction of the NBBO and trade-quote matching techniques. First, my NBBO construction approach finds the best bid and ask quotes across *all* U.S. exchanges, not only within market quotes. Second, I match trades and quotes in the same second using the interpolated time approach of Holden and Jacobsen (2014), who show that failure to perform these adjustments results in "buy/sell classification [that] are likely to be strongly biased," as mentioned above. Specifically, they show that applying the Lee and Ready's (1991) signing algorithms on whole-second TAQ without any adjustment results in 5.9% (5.7%) of sell (buy) trades that are misclassified as buy (sell) trades, and that this figure is reduced to 4.2% (5.3%) when their suggested solution is performed. The reduction in bias is substantial—sell (buy) trades that are misclassified as buy (sell) trades are reduced by 29% (7%).

This magnitude of bias reduction shown above is expected to have implications for the calculation of order imbalance, a crucial variable in my study. My analysis supports Holden and Jacobsen's (2014) finding of classification bias, since I find that imbalance measures are slightly positive when their adjustments are not implemented before signing trades. Note that this signing is for market orders; it shows only that market orders are were more likely to be sell-initiated or that traders were less patient when it came to selling compared to buying. Sell pressure declined, however, over time. For instance, the selling imbalance decreased from approximately 15% in the first sub-period to 3% in the last sub-period. It seems that there was less need in recent years for sellers to act impatiently.

Both imbalance measures (*DVOLIMB* and *NIMB*) in Table 1.1 were similar in the early years. However, a wedge developed between the two over time. In the recent subsample, the average *DVOLIMB* was -5.18% and the average *NIMB* was -4.55%; the difference is statistically significant, indicating that sell market orders tended to be larger than buy ones. This tendency for market sell orders to be large is also consistent with the aggressiveness of sell orders as compared to buy orders, at least when they dominate trading in that stock. My examination of imbalances by bin shows that their magnitude decreased in trade size over the years. The difference between *DVOLIMB* and *NIMB*, moreover, came from the small bin only.

To better understand how the size distribution of trades changed over my sample period, I examine the evolution over time of quintile thresholds of average trade size. For each month I calculate the average trade size for each stock, both in terms of dollar amount and number of shares. Then I find the average trade size in each month at the 20^{th}, 40^{th}, 60^{th}, and 80^{th} percentiles. Those thresholds are plotted in Figures 1.3A (dollar amount) and 1.3B (number of shares).

At the beginning of my sample period, the average dollar amount (number of shares) per trade at the 20^{th}, 40^{th}, 60^{th}, and 80^{th} percentiles were \$5,000 (700), \$10,000 (1,100), \$17,000 (1,500), and \$37,000 (2,200), respectively. At the end of the sample period, all quintile thresholds were below \$10,000 for dollar volume and below 500 in the case of number of shares. This result shows that in recent years more than 80% of stocks had an average trade size < \$10,000—the threshold between Bin 2 and Bin 3 in Barber et al.'s (2009) classification. This significant reduction in average trade size understates the full extent of change, given the evidence in O'Hara et al. (2014) that trades involving fewer than 100 shares were not recorded in the TAQ prior to 2014.

1.4 Conclusions

This chapter is the first in a series in this thesis that examines changes in trading activity and their implications for informed trading activity and the price impact associated with it. It outlines the data sources and explains the design of measures used in this chapter as well as subsequent chapters. In addition, it documents the fact that during the sample period 1993-2013, trading volume increased exponentially; on average, daily trading volume per stock increased from about \$2 m to \$25 m, whereas the average dollar amount per trade decreased from over \$40,000 to about \$5,000 over the same period. The proportion of small trading volume increased from 2% in early years to 42% in recent years, whereas for large trading volume the proportion dropped from 75% to 18% over the same period.

Appendix 1.A - Trade Signing Algorithms

Using order imbalance in this study requires classifying each trade as either buyer- or seller-initiated. A number of trade classification algorithms have been proposed. Three of the most important algorithms are considered here: those of Lee and Ready (LR, 1991); Ellis, Michaely, and O'Hara (EMO, 2000); and Chakrabarty, Li, Nguyen, and Van Ness (CLNV, 2007). EMO is similar to LR but its quote rule is different; a buy (sell) trade is *at* the ask (bid). CLNV also combines both quote and tick rules, and authors argue that their method is superior to those of LR and EMO. Both EMO and CLNV argue that LR approach produces biased classification for trades executed inside the quotes, and hence their approaches are geared to improve performance for these trades in particular. My main findings are based on LR due to its wide popularity[13], ensuring that my results are consistent with other studies. For robustness, I compare results based on LR with those generated using the EMO and CLNV algorithms and find that they are in agreement, though findings using LR algorithm appear to be more conservative in many instances.

Another virtue of the LR algorithm is that it leaves substantially fewer trades unclassified, especially in the early sample. I examine the proportion of unsigned trades in the early and recent samples using the three signing algorithms, both in terms of the number of trades and volume. I calculate the proportion of unsigned trades based on the number of trades for stock i at day t as follows:

$$Proportion\ of\ unsigned\ trades_{it} = \frac{\#\ of\ trades_{it} - (\#\ of\ buy\ trades_{it} + \ \#\ of\ sell\ trades_{it})}{\#\ of\ buy\ trades_{it} + \ \#\ of\ sell\ trades_{it}}$$

[13] As of November 2014, Lee and Ready (1991) have 2204 citations in GoogleScholar, versus 329 citations for Ellis, Michaely, and O'Hara (2000) and 22 for Chakrabarty, Li, Nguyen, and Van Ness (2007).

The volume-based measure is calculated as follows:

$$Proportion\ of\ unsigned\ trades_{it} = \frac{\#\ of\ shares\ traded_{it} - (\#\ of\ shares\ traded\ in\ buy\ trades_{it} + \#\ of\ shares\ traded\ in\ sell\ trades_{it})}{\#\ of\ shares\ traded\ in\ buy\ trades_{it} + \#\ of\ shares\ traded\ in\ sell\ trades_{it}}$$

Where $\#\ of\ trades_{it}$ and $\#\ of\ shares\ traded_{it}$ are daily aggregates of eligible trades in my data set. I average each of these ratios, first for each stock and then across stocks, for early (1993-2001) and recent (2002-2010) samples separately. Mean and median unsigned proportions from the second step are reported in Table 1.2.

We can clearly see that, measured in terms of both mean and median, the LR algorithm leaves far fewer trades unsigned than either the EMO or CLNV algorithms. The proportions of unsigned trades, in terms of both number of trades and volume, are lower in the recent sample. Volume-based proportions are higher than trade-based proportions, indicating that unsigned trades tend to be larger than signed trades. Pooled mean and median proportions of unsigned trades (unreported) are significantly smaller than figures presented in Table 1.2; nonetheless, I continue to use two-step statistics since they are consistent with my approach throughout this study. I also calculate but do not report those proportions using the *total* number of trades/volume in the denominator and obtain smaller figures. Therefore, my method of calculation magnifies the proportion of unsigned trades and gives an idea about the proportion of unused trades relative to my sample. The bias in order imbalance that is induced by the inability to sign some trades is not expected to be material, since it is unlikely that those trades are unsigned due to reasons that are correlated with their direction. Nevertheless, for robustness I conduct most of the tests in this analysis on a subset of stock-days where signed volume is equal to total volume and find that my findings remain unchanged. Those results are unreported, but are available upon request.

Next, the performance of the three trade signing algorithms is compared. I calculate the correlation of the daily dollar imbalance measure for each pair of algorithms, considering imbalances on both the aggregate level and in each trade size bin individually. Correlations are calculated in two steps; in the first step Pearson correlations are calculated for each stock in each sample, and then mean and median correlations across stocks in each sample are found. Results from the second step are presented in Table 1.3.

The high correlation between CLNV and EMO shows that these two algorithms are similar to each other. On the other hand, both of these measures are less correlated with LR, although their correlations with LR significantly increase in recent years. The former observation might be due to the fact that both CLNV and EMO depart from LR in their attempt to improve performance for trades executed inside quotes, whereas the latter observation may indicate that the LR rule, and its quote rule in particular, works better in assessing trades made in recent years given that there were fewer inside-quote trades post-decimalization, and in light of the abundance of quotes reported on TAQ files. In terms of individual trade size bins, correlations between CLNV and EMO are slightly higher for smaller bins, whereas correlations between LR and both CLNV and EMO are higher for larger bins. This shows that the wedge between the performance of LR and alternative algorithms widens a bit for small trades, whereas CLNV and EMO come into greater agreement for this type of trades. I confirm that my findings in this paper are not sensitive to the choice of trade signing approach, but as mentioned, I report results using imbalances based on the LR algorithm with Holden and Jacobsen’s (2014) adjustments.

Figure 1.1 - Total Volume over Time

This figure plots average daily volume per stock in each quarter, in each trade size Bin 1 (smallest) to 5 (largest), over the period 1993-2013. I follow Barber et al.'s (2009) definition of trade size bins. I first calculate average daily dollar volume in each stock-quarter by bin; cross-sectional average volume is then calculated for each quarter and stack-plotted in the figure. The two vertical lines refer to the period from the beginning of the fourth quarter of 2000 to the end of the second quarter of 2003. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped from sample if the number of daily observations is less than 50. More details on the data sample are provided in the Data section.

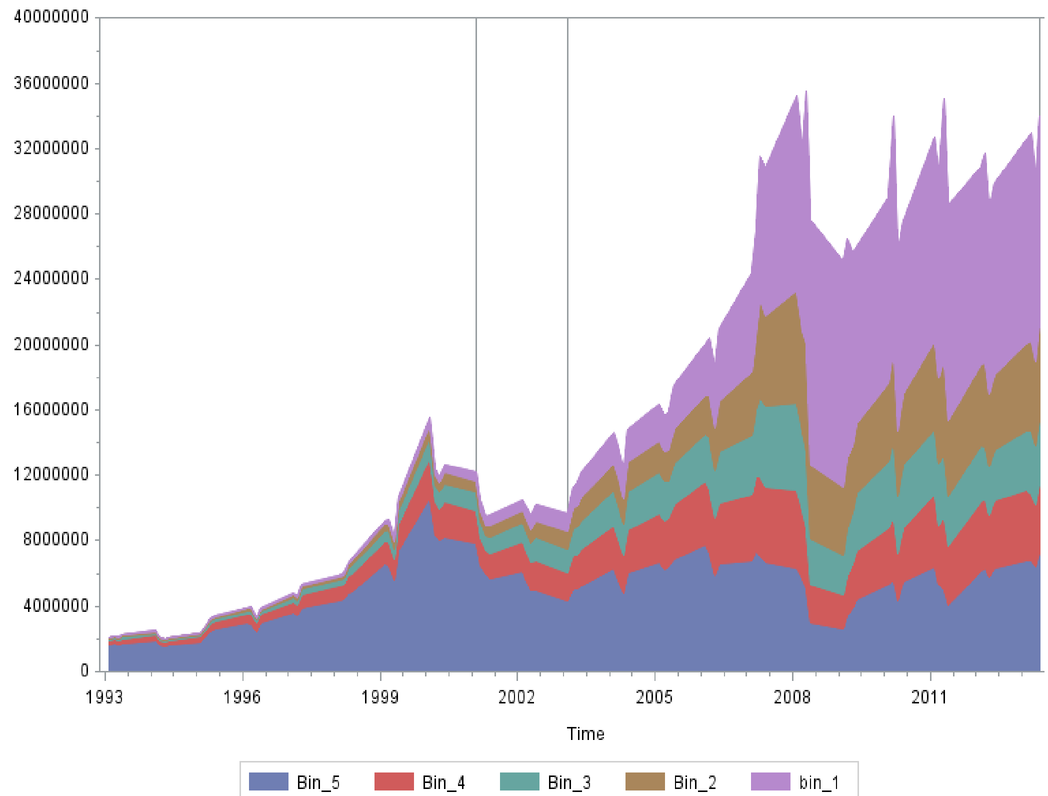


Figure 1.2 – Trade Distribution over Time

Figures plot average quarterly proportions of dollar trading volume of each trade size bin *j*=1 (smallest), …, 5 (largest), over the period 1993-2013. I follow Barber et al.'s (2009) definition of trade size bins. I first calculate bin average proportion of volume for each stock-quarter; volume-weighted and equally-weighted cross-sectional average proportions are then calculated for each quarter and plotted in the figures below. The two vertical lines refer to the period from the beginning of the fourth quarter of 2000 to the end of the second quarter of 2003. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped from sample if the number of observations is less than 50. More details on the data sample are provided in the Data section.

Figure 2A – Volume-weighted

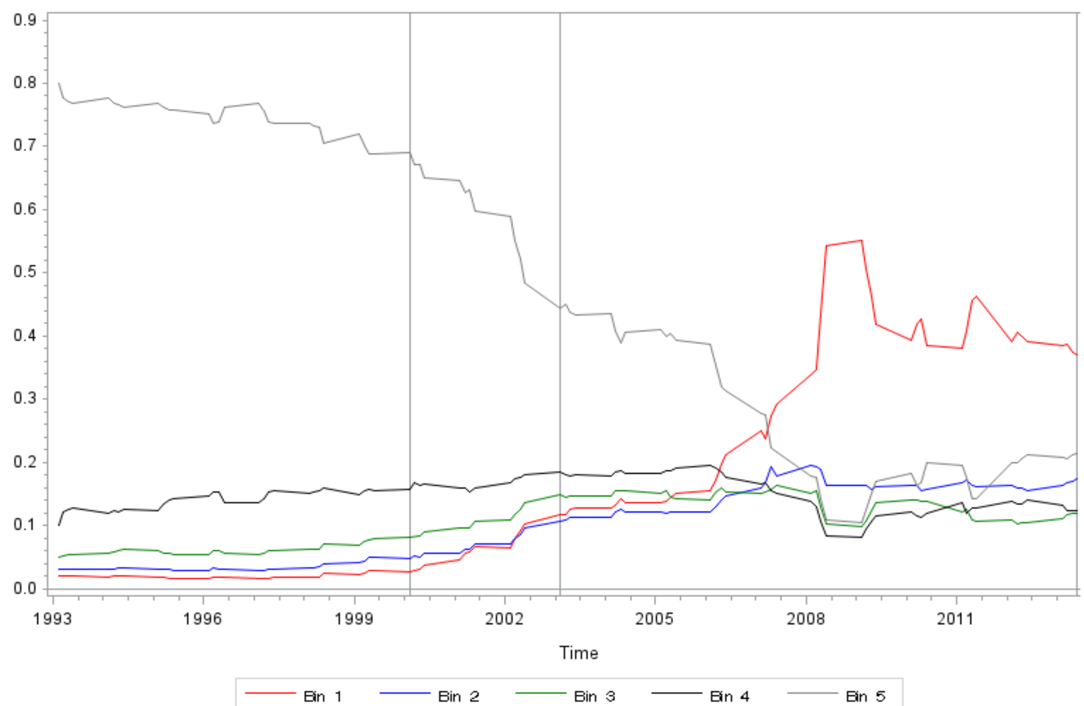


Figure 2B – Equally-weighted

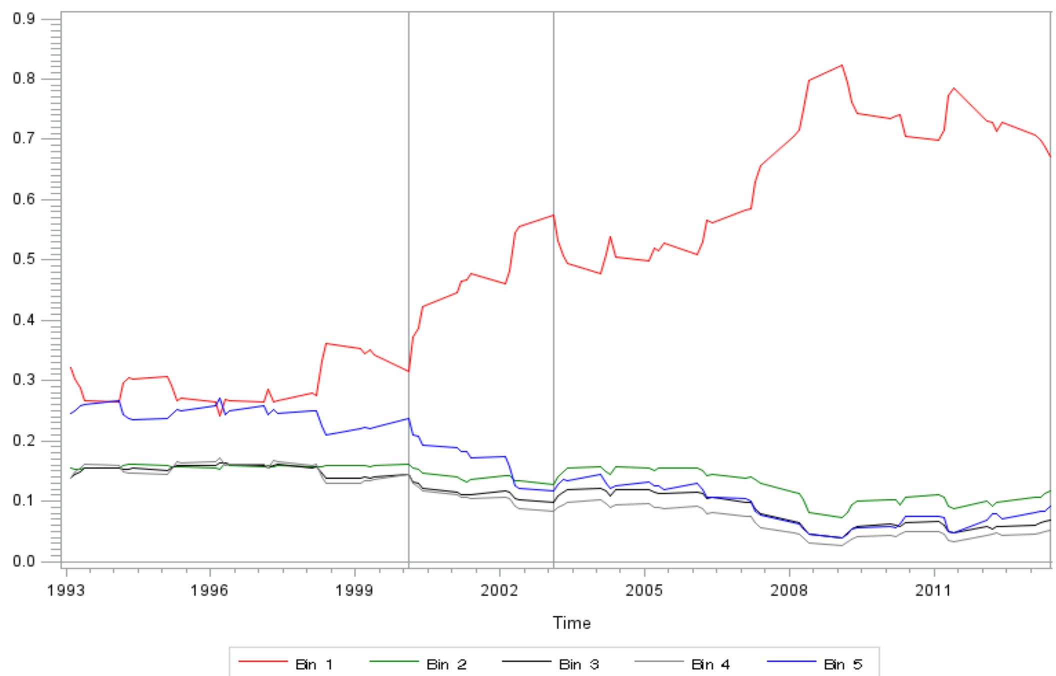


Figure 1.3 – Quintile Thresholds of Average Trade Size

Figures plot quintile thresholds of average trade size over the sample period 1993-2013, where trade size is measured by dollar amount and by number of shares. For each month I calculate average trade size for each stock, find quintile thresholds (i.e. 20th, 40th, 60th, and 80th percentiles) and plot them in graphs. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped from sample if the number of observations is less than 50. More details about data sample are provided in the Data section.

Figure 1.3A – Dollar Trade Size

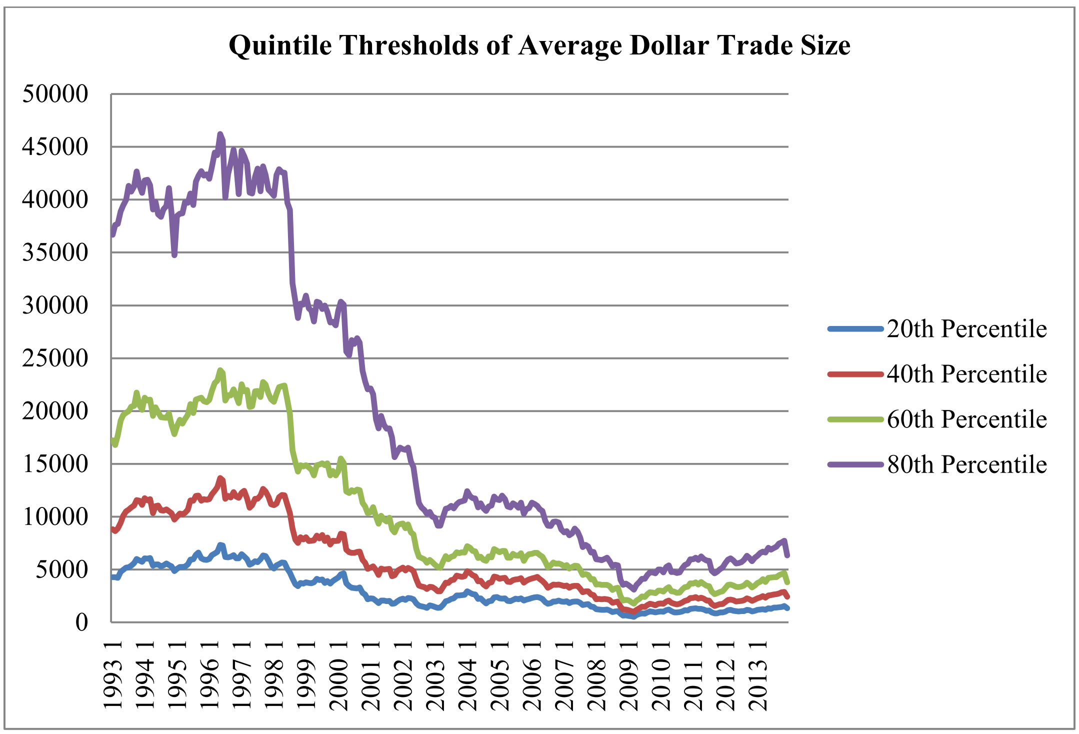


Figure 1.3B – Number of Shares per Trade

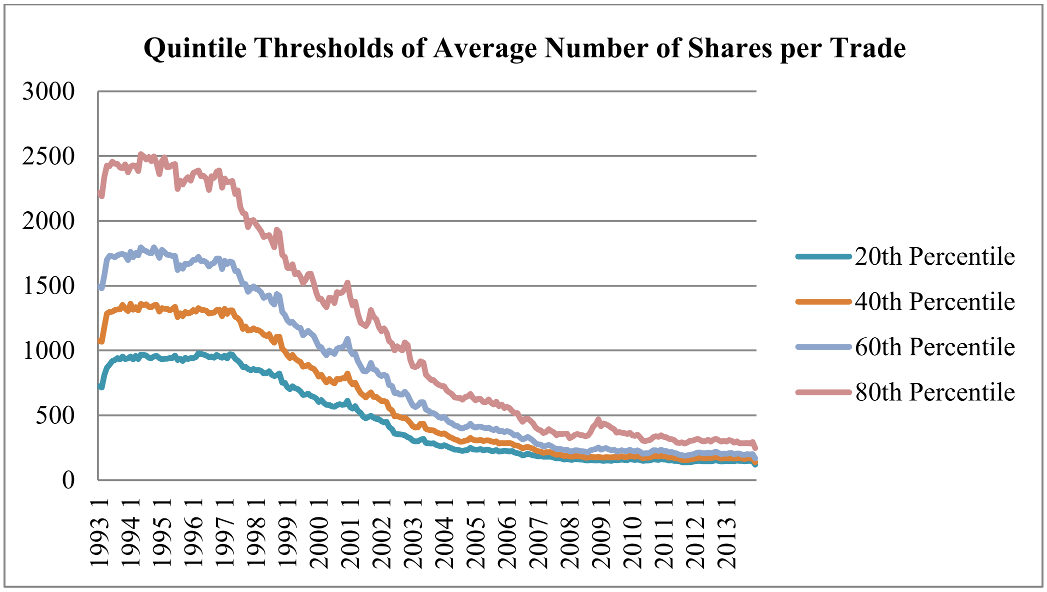


Table 1.1 - Averages of Trade Measures

This table presents averages of trade measures for stocks included in my sample over the early (1993-2001) and recent (2002-2013) samples, as well as over five subperiods. Statistics include daily averages of dollar volume (*DVOL*), number of trades (*N*), dollar volume-based imbalance (*DVOLIMB*), and number of trades-based imbalance (*NIMB*). I provide aggregate statistics and for each trade size bin *j* (*j*== 1 (smallest), …, 5 (largest)). I follow Barber et al.'s (2009) definition of trade size bins and Lee and Ready's (LR, 1991) algorithm to sign trades. Imbalance is defined as (buy-sell) / (buy+sell). The table also indicates the average proportion of volume in each bin *j*: equally- (*%DVOLj*) and volume-weighted (*%DVOLj*-W). Averages presented are cross-sectional averages of time-series averages for each stock. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped from the sample if the number of observations is less than 50. More details about data sample are provided in the Data section.

Variable	*Early*	*Recent*	*1993-1996*	*1997-2000*	*2001-2004*	*2005-2008*	*2009-2013*
DVOL	5,012,856	14,621,880	2,388,870	7,464,635	9,801,336	21,167,035	25,393,954
DVOL1	155,408	4,361,051	46,624	214,791	942,522	5,728,363	10,618,849
DVOL2	251,693	2,069,561	76,959	355,006	885,739	3,348,912	4,081,673
DVOL3	425,336	1,866,172	142,703	600,994	1,272,570	3,096,203	2,992,736
DVOL4	816,764	2,069,096	329,825	1,194,444	1,694,447	3,278,459	3,032,276
DVOL5	3,363,655	4,256,000	1,792,758	5,099,401	5,006,058	5,715,098	4,668,419
N	138	1,760	50	193	540	2,329	4,027
N1	49	1,419	15	67	320	1,771	3,454
N2	28	191	9	39	91	314	362
N3	24	88	9	34	66	147	132
N4	21	45	9	31	41	72	61
N5	15	17	8	22	21	24	17

Table 1 - Continued

Variable	*Early*	*Recent*	*1993-1996*	*1997-2000*	*2001-2004*	*2005-2008*	*2009-2013*
DVOLIMB	-0.1346	-0.0518	-0.1487	-0.1147	-0.0483	-0.0335	-0.0334
DVOLIMB1	-0.1249	-0.0457	-0.1270	-0.1139	-0.0403	-0.0286	-0.0312
DVOLIMB2	-0.0563	-0.0193	-0.0531	-0.0592	-0.0109	-0.0185	-0.0167
DVOLIMB3	-0.0408	-0.0102	-0.0389	-0.0438	-0.0003	-0.0084	-0.0131
DVOLIMB4	-0.0209	-0.0009	-0.0215	-0.0233	0.0120	0.0012	-0.0082
DVOLIMB5	-0.0055	0.0044	-0.0056	-0.0062	0.0134	0.0077	-0.0002
NIMB	-0.1345	-0.0455	-0.1412	-0.1204	-0.0430	-0.0229	-0.0275
NIMB1	-0.1290	-0.0427	-0.1295	-0.1190	-0.0394	-0.0206	-0.0266
NIMB2	-0.0564	-0.0192	-0.0531	-0.0594	-0.0108	-0.0184	-0.0165
NIMB3	-0.0408	-0.0103	-0.0390	-0.0440	-0.0004	-0.0086	-0.0130
NIMB4	-0.0212	-0.0011	-0.0216	-0.0238	0.0116	0.0010	-0.0082
NIMB5	-0.0039	0.0045	-0.0046	-0.0042	0.0151	0.0064	-0.0016
%DVOL1	32.62%	62.43%	28.67%	33.34%	52.37%	60.41%	74.20%
%DVOL2	15.95%	11.97%	15.86%	15.63%	13.57%	13.34%	9.53%
%DVOL3	14.87%	8.47%	15.73%	14.39%	10.66%	9.26%	5.61%
%DVOL4	14.53%	6.93%	15.72%	14.24%	9.36%	7.09%	4.15%
%DVOL5	22.03%	10.20%	24.02%	22.39%	14.06%	9.91%	6.51%
%DVOL1-W	3.10%	29.83%	1.95%	2.88%	9.62%	27.06%	41.82%
%DVOL2-W	5.02%	14.15%	3.22%	4.76%	9.04%	15.82%	16.07%
%DVOL3-W	8.48%	12.76%	5.97%	8.05%	12.98%	14.63%	11.79%
%DVOL4-W	16.29%	14.15%	13.81%	16.00%	17.29%	15.49%	11.94%
%DVOL5-W	67.10%	29.11%	75.05%	68.31%	51.08%	27.00%	18.38%

Table 1.2 - Proportion of Unsigned Trades

Proportion of unsigned trades in the early (1993-2001) and recent (2002-2010) samples using three trading algorithms: those of Lee and Ready (LR, 1991); Ellis, Michaely, and O'Hara (EMO, 2000); and Chakrabarty, Li, Nguyen, and Van Ness (CLNV, 2007), in terms of both the number and volume of trades. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped if the number of observations is less than 50. More details about my data sample are provided in the Data section. I calculate the proportion of unsigned trades based on the number of trades for stock *i* at day *t* as follows:

$$Proportion\ of\ unsigned\ trades_{i,t} = \frac{\#\ of\ trades_{i,t} - (\#\ of\ buy\ trades_{i,t} + \ \#\ of\ sell\ trades_{i,t})}{\#\ of\ buy\ trades_{i,t} + \ \#\ of\ sell\ trades_{i,t}}$$

The volume-based measure is calculated as follows:

$$Proportion\ of\ unsigned\ trades_{i,t} = \frac{\#\ of\ shares\ traded_{i,t} - (\#\ of\ shares\ traded\ in\ buy\ trades_{i,t} + \ \#\ of\ shares\ traded\ in\ sell\ trades_{i,t})}{\#\ of\ shares\ traded\ in\ buy\ trades_{i,t} + \ \#\ of\ shares\ traded\ in\ sell\ trades_{i,t}}$$

Where $\#\ of\ trades_{i,t}$ and $\#\ of\ shares\ traded_{i,t}$ are daily aggregates of eligible trades in my data set. In the case of each of these ratios, I average them first for each stock in the early and recent sample, and second across stocks, for early and recent samples separately. Mean and median (in brackets) unsigned proportions from the second step are reported in the table.

Algorithm	Number of trades		Volume	
	Early	Recent	Early	Recent
LR	0.019 (0.011)	0.002 (0.0005)	0.035 (0.013)	0.006 (0.001)
EMO	0.188 (0.131)	0.049 (0.010)	0.328 (0.195)	0.112 (0.029)
CLNV	0.171 (0.121)	0.024 (0.006)	0.298 (0.179)	0.061 (0.017)

Table 1.3 - Correlations between Trade Imbalances

Mean correlation between imbalances based on the three trade signing algorithms used: those of Lee and Ready (LR, 1991); Ellis, Michaely, and O'Hara (EOH, 2000); and Chakrabarty, Li, Nguyen, and Van Ness (CLNV, 2007). When trades across all trade sizes are considered, imbalance of stock i at day t ($DVOLIMB_{it}$) is calculated as follows: $\frac{DVOLBUY_{it} - DVOLSELL_{it}}{DVOLBUY_{it} + DVOLSELL_{it}}$, where $DVOLBUY_{it}$ ($DVOLSELL_{it}$) is the aggregate Dollar value of buy (sell)-initiated trades of stock i at day t. When trades in trade size bin j is considered, where j = 1, …, 5, imbalance of stock i at day t ($DVOLIMBj_{it}$) is calculated as follows: $\frac{DVOLBUYj_{it} - DVOLSELLj_{it}}{DVOLBUYj_{it} + DVOLSELLj_{it}}$, where $DVOLBUYj_{it}$ ($DVOLSELLj_{it}$) is the aggregate Dollar value of buy (sell)-initiated trades in size bin j of stock i at day t. I follow Barber et al.'s (2009) definition of trade size bins. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped if the number of observations is less than 50. More details about my data sample are provided in the Data section. I find correlations between pairs of imbalances based on the three trade signing algorithms I use: LR, EOH, and CLNV. I first find correlations for each stock in the early (1993-2001) and recent (2002-2010) samples, and then report the equally-weighted cross-sectional average correlations across stocks in sample.

	Pair of Algorithms	***Early***	***Recent***
All trade sizes	*CLNV/EOH*	0.961	0.894
	LR/CLNV	0.616	0.918
	LR/EOH	0.576	0.813
Bin 1	*CLNV/EOH*	0.987	0.919
	LR/CLNV	0.564	0.953
	LR/EOH	0.613	0.865
Bin 2	*CLNV/EOH*	0.969	0.872
	LR/CLNV	0.683	0.918
	LR/EOH	0.649	0.790
Bin 3	*CLNV/EOH*	0.966	0.876
	LR/CLNV	0.696	0.891
	LR/EOH	0.661	0.773
Bin 4	*CLNV/EOH*	0.970	0.885
	LR/CLNV	0.715	0.874
	LR/EOH	0.666	0.750
Bin 5	*CLNV/EOH*	0.953	0.879
	LR/CLNV	0.716	0.840
	LR/EOH	0.669	0.729

Chapter 2 - Implications of Changes in Trading Activity for Price Impact Patterns of Trades

Introduction

The first chapter of this thesis quantifies changes in the proportions of each trade size. This finding suggests the need for a second study that focuses on price formation over the sample period, contrasting the permanent and transitory price effects of trades conditional on their sizes, and other non-trade information. Market microstructure theory shows that price changes can be modeled as a function of order flow because order flow conveys information (Kyle [1985], among others). The presence of trading frictions in practice impedes the incorporation of information into prices (Chordia et al., 2011). If recent changes in markets have affected these frictions, then reexamining the trade-price relation is of great importance.

In this chapter, I study both the permanent information content of trades and temporary price pressure, examining how these effects varied by trade size and over the sample period. On one hand, using Easley and O'Hara's (1987) theory leads to the prediction that market is in a pooling equilibrium currently (i.e. informed and uninformed traders are less separated). This prediction implies that informed traders followed the shift in the distribution of trades. On the other hand, the literature shows that informed traders find it infeasible to trade in small sizes (e.g. Barclay and Warner, 1993) because of their need to reach their desired position as quickly as possible before their information becomes stale. Therefore, if the increased proportion of small trade that I document is primarily originating from, for instance, small retail traders who currently face lower barriers to participating in markets, I might not find that the price impact of small trades has increased, because retail traders are typically considered less-informed and hence their contribution to permanent price changes is limited. In this case, even if the collective large volume of small trades exerts price pressure due to the liquidity effect, price changes will eventually be reversed, leaving a minimal permanent price impact.

In the study that follows, I document important shifts in price impact patterns. My results show that medium trades were associated with most of the price changes during the earlier years, but that in later years, small trades exert most of the permanent price pressure. This more recent small trade influence indicates that, in order to remain undiscovered, informed traders have indeed moved towards small sizes to match the new distribution of trades, as theoretical microstructure models predict. My findings have direct implications for the stealth trading literature. The stealth trading hypothesis posits that traders with private information trade gradually and in non-large sizes to avoid market attention. Barclay and Warner (1993) find that those traders concentrate their trades on medium-sized transactions. My data suggest that, in the early years of my sample, medium trades had incremental power for price impact, above that implied by their trading volume share. This incremental explanatory power, however, moved to small trades in the later years of my study. This change shows the growing importance of small trades in recent years; it suggests that they warrant additional attention from academics and professionals. I examine the nature of this shift in trade size distribution and in the process of price discovery. I find that most of the volume shift towards small trades originated from medium bins. The increase in small trade price impact also comes primarily at the expense of medium trades.

The remainder of this chapter is organized as follows. Section 2.1 reviews relevant streams of the literature and develops the main hypothesis. Section 2.2 presents my main tests concerning the determinants of order imbalance and the price impact of trades. The section also relates my findings to the context of the stealth trading literature. Section 2.3 examines the role of public information versus that of private information in price discovery. Section 2.4 concludes the chapter.

2.1 Literature Review

In this section I discuss the pertinent literature and highlight its role in guiding my investigation. My central hypothesis is that the size distribution of trades has shifted toward a greater proportion of small transactions, and that this shift was accompanied by a similar shift in the price impact of trades. In the following subsections, I review findings from relevant streams of the literature.

2.1.1 Stealth Trading

Trades in general, and large trades in particular, are associated with price impact. There are at least two reasons why prices respond to trades. The first is the liquidity effect: large trades tend to move inventories away from optimal levels. This effect, however, is temporary, as it usually reverses after order flow stabilizes and inventories are adjusted (Subrahmanyam, 2008). The second reason is the adverse selection problem, which emerges as a consequence of buyers and sellers trading against informed traders, resulting in a permanent price impact. The second of these reasons is more closely related to my analysis.

A number of studies have shown that informed traders have an incentive to trade in large sizes in order to maximize return on information they possess. In such situations, market makers' strategies for setting prices depend on trade size. Easley and O'Hara (1987) find that price concessions increase with trade size, and that there is a positive relation between the size of a trade and the probability that the trade is information-based. They also describe two types of equilibria in markets: separating equilibrium and pooling equilibrium. A separating equilibrium takes place when informed traders trade in large quantities only and are *separated* from the rest of the market, whereas a pooling equilibrium occurs when informed traders trade in both small and large sizes. The theoretical work of Easley and O'Hara (1987) shows that the market will be in a separating equilibrium if it has sufficient width (the ratio of large to small trades), or if there are few information-based trades. Alternatively, the market will be in a pooling equilibrium if it is narrow or shallow, or if there are many information-based trades (proposition 3 of their paper). The findings of subsequent empirical research, however, depart from the theoretical framework of Easley and O'Hara (1987). Barclay and Warner (1993) show empirically that medium-sized trades are associated with the largest price impact. Their findings were later confirmed by Chakravarty (2001). Barclay and Warner (1993) argue that, in theory, informed traders opt for trading in large sizes[14], but this behavior exposes the nature of their trades to the market, so they tend to split their trades. However, the offsetting cost of trading in smaller sizes is a delay in the acquisition of desired positions because the processing of trades does not occur quickly enough. This delay might cause prices to move against those investors before they acquire their positions if information becomes more readily available in the meantime. In addition, the structure of trading costs makes it very expensive to split large trades into many small trades.

Balancing the advantages and disadvantages of large and small trade size strategies, informed traders decide to trade in medium-sized transactions; this leads to a certain degree of pooling in markets. Therefore, medium-sized trades witness a concentration of information-based trading and are expected to be associated with the largest price impact. Barclay and Warner (1993) find

[14] This prediction has its roots in the theoretical work of Kyle (1985). Kyle shows that under certain assumptions, informed traders trade as much as required for prices to reach their full with-information values.

evidence in favor of this hypothesis, which is consistent with a stealth trading explanation. To summarize, these studies of empirical stealth trading show that while informed traders try their best to hide their identity and the nature of their trades, they avoid full integration with the typically-uninformed small traders. That avoidance is due to the following three reasons: *markets for large and medium trades are not shallow or narrow* (Easley and O'Hara, 1987), *trades are not processed quickly enough*, and *small trades are cost-ineffective* (Barclay and Warner, 1993).

The years subsequent to those studies witnessed significant regulatory and technological changes in stock markets, and evidence shows that those changes have had direct implications for the above three issues in particular. For instance, Goldstein and Kavajecz (2000) and Chakravarty et al. (2005) found that depth declined after tick reduction in 1997 and decimalization in 2001, respectively. Those findings, assessed from the perspective of Easley and O'Hara's (1987) framework, suggest that the market for non-small trades has become shallower; it has therefore become even harder for informed traders to continue trading in large or medium quantities. In terms of the speed of trade processing, the implementation of a number of systems supporting the automatic execution of trades in U.S. markets, such as Direct+ (2000) and Hybrid (2006) on the NYSE, has led to significant reductions in trade processing time (Jain [2005], and Hendershott and Moulton [2011]). The implementation of the Direct+ system is of particular relevance to my hypothesis. Direct+ is the first system to offer automatic *execution* of trades, but only for relatively small trades involving fewer than 1100 shares.

Finally, new advances in stock markets have reduced trading costs significantly (Chordia et al., 2011). Most of the studies that examine the change in cost around events demonstrate that the cost has decreased in general. However, some studies provide evidence that the cost reduction is more pronounced for small trades. In fact, Goldstein and Kavajecz (2000) show that the trading costs following the 1997 tick-reduction event have *increased* for large trades and that a cost reduction is documented for smaller trades only. Chakravarty et al. (2005), examining the effect of the 2001 decimalization event on trading costs, produce results that echo those reported by Goldstein and Kavajecz (2000). They find that costs increased for orders that aggressively sought liquidity (not-worked orders; those not filled within a day), but declined for worked orders. They conclude that decimalization appears to have benefited institutions that were working their orders on the exchange floor.

The studies by Goldstein and Kavajecz (2000) and Chakravarty et al. (2005), in short, showed clearly that the cost savings advantage generated by the two tick-reduction events is limited to small trades, giving informed traders further incentive to trade in small sizes. Tick reduction events, moreover, have an additional effect on the trading strategies of informed investors. Harris (1996) shows that the trade size is positively related to tick size. The rationale is that the cost of front-running declines as tick size is reduced. Therefore, concerned about being front-run, block traders are expected to avoid exposing their entire trades at once after the 1997 and 2001 tick reductions.

Facing all of the above changes in the market, informed traders benefit from small size trades in two ways: directly and indirectly. The direct benefit, shared with all traders, is related to reduced cost and increased speed, while the indirect benefit, exclusive to informed traders, stems from the fact that having a larger group of small uninformed traders with whom to mingle allows for more effective stealth trading. Therefore, I argue that informed traders strategically choose small trades as a new optimal size decision in the context of the tradeoff problem described by Barclay and Warner (1993). My main hypothesis is as follows:

The proportion of small trades increased, and the informativeness by information-based trades shifted from the medium-sized to small trade category, from the early to the recent sample periods.

Various changes in markets might have contributed to this conjectured phenomenon. However, I split my sample roughly around the year 2001, expecting to capture this shift, since it was around that time that markets witnessed many changes with implications for both depth and speed, as discussed in this paper[15]. Finding support for this hypothesis challenges both common belief and the evidence in the literature that the highest price impact is caused by large (Chan and Fong, 2000) or medium trades (Barclay and Warner [1993], and Chakravarty [2001]).

2.1.2 Transparency

My analysis is also related to the literature about transparency in stock markets. Various theoretical and empirical papers make predictions about the relationship between the state of market transparency and trading strategies. Keim and Madhavan (1996) show that knowledge of the identity of traders affects the price impact of trades and that there is a lack of anonymity for large trades. Evidence shows that there is a negative relationship between transparency and trade size.

In January 2002, the NYSE launched the OpenBook system, which provides limit order information off-exchange. Boehmer et al. (2005), who studied this event, found that traders tended to submit smaller trades after the implementation of the system. Such transparency-increasing systems make it harder for informed traders to trade in large or even medium

[15] More information about NYSE events can be found on the NYSE website: http://www1.nyse.com/about/history/timeline_regulation.html.

quantities, assuming they choose limit orders. The anonymity of information-based trades is at higher risk of being exposed to the market unless informed traders adapt to the new market environment by mixing with smaller investors, especially in light of the declining costs of small trades, and by trading electronically, thereby benefiting from the anonymous nature of electronic trading platforms.

2.2 Tests and Results

2.2.1 Dynamics of Order Imbalance

To learn about the dynamics of trading activity, I regress daily imbalances on a lagged daily imbalance, controlling for lagged own and market returns. I estimate the following two regression equations:

$$DVOLIMB_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \beta'_{it}\, MKTRT_{t-1} + \gamma_{it}\, DVOLIMB_{it-1} + e_{it} \quad (1)$$

$$DVOLIMBj_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \beta'_{it}\, MKTRT_{t-1} + \sum_{j=1}^{5} \gamma_{ijt-1} DVOLIMBj_{it-1} + e_{it} \quad (2)$$

This test helps to answer the following questions: (i) What are the cross- and auto-correlations in imbalances and how have they changed over time? These correlations shed light on the extent and trend in order splitting phenomenon in recent years. (ii) How strong is return-chasing trading, then and now? (iii) What are the R-squared values of those regressions? R-squared values determine the explanatory power of order imbalance.

I estimate this regression, as well as the rest of the regressions, by first estimating the equation for each stock, and then providing cross-sectional and statistical significance based on the t-stat of the average[16]. Standard errors are corrected for cross-correlations using the assumptions provided in Chordia and Subrahmanyam's (2004) study, in which residual correlation is proxied by the average correlation calculated from groups of 100 stocks separated based on their PERMNOs. A number of similar studies report cross-sectional statistics concerning coefficients obtained from time-series regressions. For instance, Chan and Fong (2000) and Chordia and Subrahmanyam (2004) report the cross-sectional average coefficients obtained from time-series regressions, whereas Harford and Kaul (2005) report the cross-sectional median coefficients. To keep my results as close as possible to those studies, I follow their regression style, which involves providing cross-sectional statistics of time-series regression estimates. Alternative estimation methods such as using Fama-MacBeth's (1973) regressions or clustering standard errors by date following the recommendation of Petersen (2009) yield similar results[17].

Results are presented in Table 2.1. Average coefficients on lagged aggregate imbalance (referred to here as *auto-correlations*) are consistently positive in both periods but declined from 0.21 in the early sample to 0.15 in the recent sample; this difference is statistically significant. The lower auto-correlation may reflect the fact that technologically-advanced and more liquid markets allow for faster trading; hence more trades are executed *within* the day even if a trade splitting strategy is followed. If it is becoming easier to process most or all child trades within one day even though more splitting is occurring, an examination of *intra-daily* autocorrelation is

[16] In all of my regressions, in addition to assessing mean coefficients, I also consider but do not report median coefficients, the proportion of coefficients that are statistically significant, and the proportions of significant coefficients that are positive/negative. My findings are based on all of these statistics.

[17] Petersen (2009) shows that clustering standard errors by the time dimension only or conducting Fama-MacBeth (1973) regressions is appropriate in asset pricing regressions, since stock market data usually suffer from time effect but not firm effect.

likely to reveal this. Imbalances in bins also show less autocorrelation in recent years, but the extent of autocorrelation reduction increases with bin size. Increased splitting in small bins seems to counteract the general trend of autocorrelation reduction. Autocorrelations become insignificant for Bins 3-5.

Average coefficients on own lagged return shifted from negative values in the early sample to positive values in the recent sample. Investors, that is, tended to sell winners and buy losers on the following day in the early years, but did the opposite in recent years. Average coefficients for individual bins in recent years are insignificantly different from zero. Average coefficients on lagged value-weighted market return also generally decreased between the two periods, and the decrease was more pronounced for larger bins.

In the next section I estimate regression models in which return is on the left-hand side of the equation and lagged return, contemporaneous imbalances, and lagged imbalances are on the right-hand side. Because imbalances in different trade size bins can move together, it is important to ensure that including imbalances in all bins together in the regression equation does not result in multicollinearity. I calculate the VIF statistic and find it to be less than 2 for all variables in the different regression models. I also examine correlation coefficients among imbalance variables. The coefficients of lagged imbalances shown in Table 2.1 can be treated as correlations between contemporaneous and lagged imbalances (regression correlations). In addition, I calculate correlation coefficients between contemporaneous order imbalance variables.

Table 2.1 presents coefficients of lagged imbalances included in the regressions explained above. These results reveal that correlations between contemporaneous and lagged imbalances in different bins are generally low—below 0.2. Cross-sectional average correlation coefficients among contemporaneous imbalances in different bins (unreported) are not severely high; they range between 0.11 and 0.36 in the early sample, and between 0.03 and 0.22 in the recent sample. Considering that VIF statistics are generally small, these moderate correlations alleviate concerns about any multicollinearity issues in subsequent regression analyses. In addition, while correlations and regression correlations are generally positive, they seem to decrease as the distance between bins increases, and they are generally lower among larger bins than among smaller bins. This pattern is consistent with that found by Barber et al. (2009). These results show that net trades in different bins, while generally in the same direction across all bins, are more closely related in the case of small trades. There seems to be more coordination among small trades than among large trades.

2.2.2 Price Informativeness of Trades

The previous chapter sheds light on the extent of shift in trade distribution and offered preliminary evidence on the increasing informativeness of small trades. This subsection formally tests whether informed traders moved to match the new distribution of trades, by examining the price impact of imbalances and its change over time. The test involves estimating a regression equation in which stock return is located on the left-hand side of the equation, and contemporaneous and lagged imbalances ($DVOLIMB$) are placed on the right-hand side. Results based on *NIMB* are qualitatively similar so they are unreported. In one model, I consider aggregate imbalance, and in another, I consider imbalances by bin. The two models are as follows:

$$RETX_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \sum_{T=\,t,t-1} \gamma_{iT}\, DVOLIMB_{it} + e_{it} \qquad (3)$$

$$RETX_{it} = \propto_{it} + \beta_{it} RETX_{it-1} + \sum_{j=1}^{5} \gamma_{ijt} DVOLIMBj_{it} + \sum_{j=1}^{5} \gamma_{ijt-1} DVOLIMBj_{it-1} + e_{it} \qquad (4)$$

I refer to the estimate of imbalance coefficients (γ) by price impact. Note that I also include lagged *RETX*. Chordia and Subrahmanyam (2004) do not include lagged return in their regression model since it could be collinear with imbalance. I find that controlling for lagged return in my model does not affect other coefficients. Chordia and Subrahmanyam (2004) also include five lags of imbalances. I found that imbalances beyond the first lag are mostly statistically insignificant.

Panel A of Table 2.2 presents the estimation results of equation 3, in which standard errors are corrected as in Models 1 and 2. Figure 2.1 plots quarterly estimates of contemporaneous (γ_t) and lagged (γ_{t-1}) price impacts from equation 2 as well as the combined effect ($\gamma_t + \gamma_{t-1}$). There is no discernable pattern in price impact over time. Table 2.2 shows a slight increase for price impact in the recent sample but, as Figure 2.1 shows, most of the difference is related to the period around the 2008 financial crisis. To express price impact in economic terms, one standard deviation of increase in daily imbalance is associated with an approximately 0.5% increase in the same-day return in the early sample and a 0.7% increase in the recent sample.

Results also show, as expected, that lagged imbalances are negatively related to returns, as mentioned above. The coefficient on lagged imbalance is usually interpreted as reversal related to liquidity effect. Results in Table 2.1 show that contemporaneous and lagged imbalances are positively correlated. These seemingly contradictory findings are consistent with the results documented by Chordia and Subrahmanyam (2004), who offer an interesting explanation for the apparent conflict. Serial correlations in imbalances mean that imbalances can be decomposed into two components: one that is history-dependent and one that is new. Regressions weight the two components in the contemporaneous imbalances equally, but the new component is much more relevant to returns than the history-dependent one, resulting in over-weighting for the latter. This over-weighting is counteracted by coefficients with opposite signs associated with lagged imbalances.

Panel B of Table 2.2 presents estimation results of equation 4. Figure 2.2 plots combined quarterly estimates of price impact ($\gamma_t + \gamma_{t-1}$) for each bin. The figure shows that in the early years of the period studied (1993-1997), the price impact of imbalances in different trade-size bins were bounded by a tight range, with the price impact of medium-sized trades at the top, followed by large trades, and then small trades with the least impact.

Results suggest that, starting in the middle of 1997, this pattern changed completely as the small bin began to have the highest impact and the large bin the smallest impact. Estimates in Panel B of Table 2.2 show that changes in the price impact of imbalance are strongly negatively correlated with the size of trades. The impact of Bin 1 doubled from 0.021 to 0.0425; the change for Bins 2 and 3 is statistically insignificant; the impact for Bin 4 decreased slightly, from 0.0066 to 0.0046; and the impact for Bin 5 demonstrated the most significant decrease, from 0.0039 to 0.0013. Price impact trajectories in Figure 2.2 resemble to a large extent those of the trade distribution in Figure 1.2. As mentioned, Figure 2.2 plots the sum of contemporaneous and lagged price impact. Regression results here show that the latter is negative; this is interpreted as reversal due to price pressure and temporary inventory effects. Therefore, the net price impact in Figure 2.2 is mainly information-driven. Those changes in price impact patterns indicate that the shift in trade distribution towards smaller trades is accompanied by a migration of informed traders in the same direction, causing a similar shift in the price impact.

Within the framework of Barclay and Warner's (1993) argument, this finding shows that the most recent years, in which I observe a shift towards a larger price impact for small trades, witnessed some changes in markets that allowed informed traders to split trades while protecting their informational advantage, thereby reducing the second offsetting factor explained above. The significant drop in the price impact of largest trades in recent years may be explained on the grounds that trades in this category are dominated by transactions, such as dark pool trades, which were negotiated (during which negotiations specialists and market makers were assured that they were not information-driven).

The diminishing average trade size and the increasing impact of small trades imply that, in aggregate, trade measures based on the number of trades have become more effective in explaining returns than measures based on trading volume. This conjecture is based on the insight that measures based on the number of trades vary more with small trades than their volume-based measures do. I test this implication by including the two imbalance measures (*DVOLIMB* and *NIMB*) in equations 3 and 4. In aggregate model 3 (results are unreported), *NIMB* dominates *DVOLIMB*; the latter becomes only marginally significant in the early sample and completely insignificant in the recent sample. *NIMB* does not gain significance over *DVOLIMB* when imbalances by bin are included. In addition I replace *DVOLIMB* by *NIMB* in model 3 and plot quarterly combined price impacts from the two variations of the model in Figure 2.3. In the early years, prior to 1997, the price impacts of the two measures were not statistically different and almost identical in magnitude, but since 1997, the price impact of *NIMB* clearly exceeded that of *DVOLIMB*[18].

2.2.3 Stealth Trading

As mentioned above, my finding that the imbalance in medium-sized trades is associated with the highest price impact in early years is consistent with what a number of studies, such as those of Barclay and Warner (1993) and Chakravarty (2001), have documented. Both papers concluded that the disproportionately large price impact for medium-sized trades supports the stealth trading hypothesis. My findings show that this disproportionately large impact moved to small trades and indicate that stealth trading became concentrated in the small bin in recent years. Note that my imbalance measures control for volume in the denominator, so the price impact associated with this measure can be interpreted as an incremental effect of the proportional imbalance along with the effect of volume. As an alternative way to assess stealth trading, I regress return on lagged return, and on contemporaneous and lagged imbalances, both in the aggregate and by bin, in the same equation. The sums of coefficients on contemporaneous and lagged imbalances by bin represent the incremental impact of trading. Those are plotted in Figure 2.4A. This figure shows that the maximum incremental effect occurred in medium size bins in the early years, but that the incremental effect in Bin 1 began to dominate a few years after the initial (starting) date of the sample.

To render my results in a form more easily interpreted as evidence regarding the stealth trading hypothesis, I compare in each bin the proportion of volume versus the proportion of price

[18] Jones et al. (1994) find that the number of trades is the main factor behind the volume-volatility relationship, and that neither trade size nor order imbalance has additional explanatory power. My findings are partially consistent with those in Jones et al., but only in the recent years of my sample period.

impact (combined coefficients in each bin divided by the sum of combined coefficients).[19] If a certain trade-size bin has a larger share of the price impact than that of its volume, this is evidence of concentrated informed trading. I calculate for each stock in each quarter the average difference between the price impact and volume proportions and then average that value across stocks in each quarter. Results that are plotted in Figure 2.4B are generally in agreement with those in Figure 2.4A. Specifically, the difference between the proportion of price impact and the proportion of volume was the highest in the medium-sized Bin 3 in the early years of the sample, but after those years the difference became highest in Bin 1. Interestingly, the sum of differences across bins decreased around the time of the 2008 financial crisis, indicating that total information asymmetry is reduced during turbulent market periods.

2.3 Public versus Private Information

Price changes can result from private information being conveyed through trades or from public information such as news announcements that become incorporated into prices without the need for trading. In the latter case, quotes get revised once public information is disseminated. In Hasbrouck's model (1991), the R-squared of the return regression on trading variables and the complement of R-squared represent the roles of private information and non-private information, respectively. Table 2.2 shows those average R-squared values for regression Models 3 and 4. The average R-squared values for Model 3 of the aggregate imbalance changed only slightly between the early years and the recent years, and the change was statistically insignificant. On the other hand, the R-squared values for Model 4, which measures imbalances by bin, jumped by

[19] This is just a rough test. In their study, Barclay and Warner (1993) compare the proportion of price changes in each trade-size category with the proportion of volume in that category. The data I am using for this study are aggregated at the daily level, and my test is conducted in the spirit of their idea.

about 40%, from 5.28% to 7.11%. In the early years, moreover, average R-squared values derived from both models were not significantly different from each other (5.38% versus 5.28%), while they were significantly different in the more recent sample (5.76% versus 7.11%). This divergence between the two models in recent years suggests that during that period, unlike in the early years, distinguishing between small and non-small trades allowed traders to learn more information. This probably indicates that informed trading is less biased by size (towards larger-scale transactions) in more recent years than previously.

Next, I added imbalances in individual bins to the model sequentially in different orders. Average R-squared values for all combinations of variables are presented in Table 2.3. This sequential addition of imbalances helped me to learn more about the boost to the explanatory power of the model and about which bins in particular drive this increase. When lagged return is included alone, R-squared drops significantly in the recent sample. Serial correlations in return are less in recent years, a finding consistent with greater efficiency in markets (Chordia, 2011). Looking at R-squared values for the rest of the models, the model with Bin 3 has the highest average R-squared value in early years, closely followed by the Bin 1 model. This is consistent with early evidence that medium-sized trades are the ones with the highest probability of information. An examination of average R-squared values in recent years shows that the R-squared measure for the Bin 1 model is the only one that experienced an increase, whereas values for all other models with any of the other bins individually decreased. In addition, when Bin 1 is added to any other combination of bins in the model, R-squared value increases significantly, but not vice versa.

2.4 Conclusions

Equity markets witnessed significant changes in recent years. Evidence in Chapter 1 shows that trading volume increased and average trade size decreased. Markets have lower depth than they previously had, and trading became faster, easier, and less expensive. These market conditions facilitated trading in small sizes for all traders. Informed traders have an additional incentive to trade in small quantities. They are attracted to the large crowd of small traders because the large number of small trades offers them an anonymous environment for trading in an increasingly transparent market. Consistent with this conjecture I offer evidence that volume distribution generally shifted over the period of the study towards smaller transactions, and that this shift was dominated by informed traders who quickly adapted to the new market conditions. This conclusion is evident in the substantially larger share of small trades in price discovery, which may be attributed to informed traders. This study is a vivid example of how findings in previous market microstructure literature might have changed drastically in response to changes in markets that occurred in recent years.

Figure 2.1 – Price Impact of Trades over Time

This figure plots quarterly price impact of dollar volume imbalance (*DVOLIMB*) over the period 1993-2013. I follow Lee and Ready's (LR, 1991) algorithm to sign trades. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped from the sample if the number of observations is less than 50. More details about data sample are provided in the Data section in Chapter 1. To estimate price impact I run the following regression for each stock using daily observations over a quarter:

$$RETX_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \sum_{T=t,t-1} \gamma_{iT}\, DVOLIMB_{it} + e_{it}$$

Where $RETX_{it}$ is stock's *i* return at day *t*. I find γ_{it}, γ_{it-1} and the sum of the two (Total) for each stock, and plot the cross-sectional averages of the series for each quarter. The two vertical lines refer to the period from the beginning of the fourth quarter of 2000 to the end of the second quarter of 2003.

Figure 2.2 – Price Impact of Trades over Time (By Bin)

This figure plots quarterly price impact of dollar volume imbalances for each trade size bin *j* (*DVOLIMBj*), where *j* = 1 (smallest), …, 5 (largest), over the period 1993-2013. I follow Barber et al.'s (2009) definition of trade size bins and Lee and Ready's (LR, 1991) algorithm to sign trades. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped from the sample if the number of observations is less than 50. More details about data sample are provided in the Data section in Chapter 1. To estimate price impact I run the following regression for each stock using daily observations over a quarter:

$$RETX_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \sum_{j=1}^{5} \gamma_{ijt}\, DVOLIMBj_{it} + \sum_{j=1}^{5} \gamma_{ijt-1} DVOLIMBj_{it-1} + e_{it}$$

Where $RETX_{it}$ is stock's *i* return at day *t*. I sum γ_{ijt} and γ_{ijt-1} for each stock, and plot the cross-sectional average for each bin in each quarter. The two vertical lines refer to the period from the beginning of the fourth quarter of 2000 to the end of the second quarter of 2003.

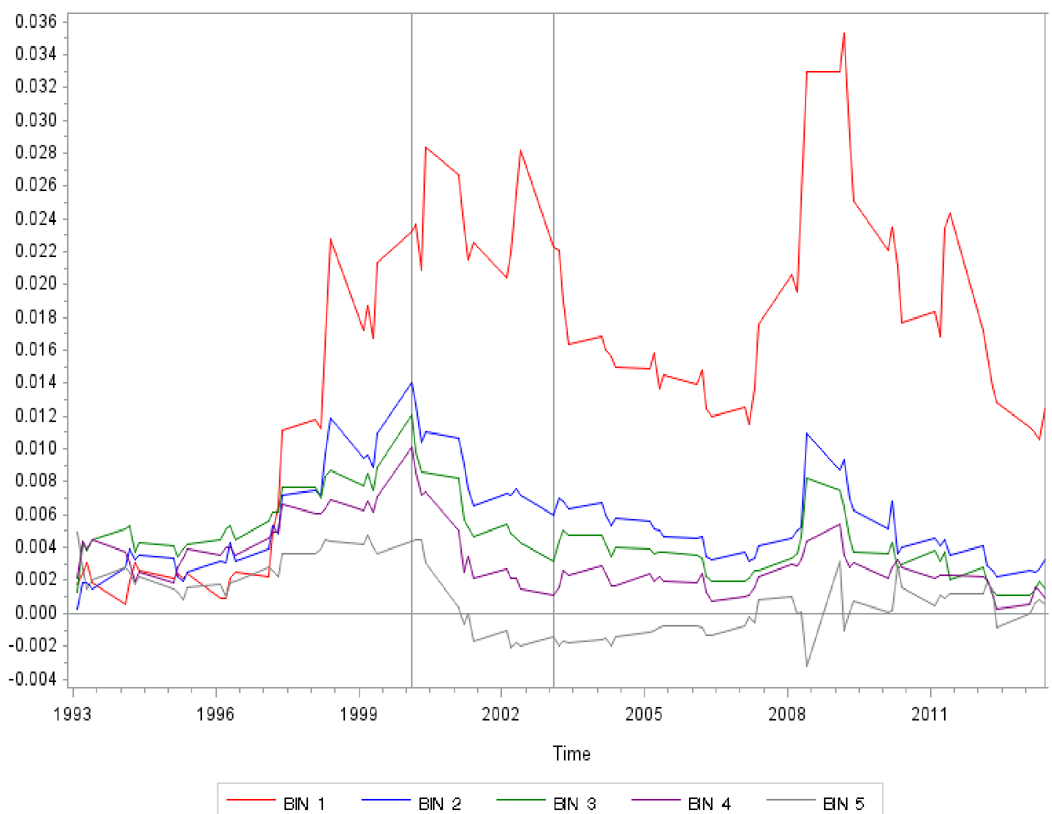


Figure 2.3 – Price Impact of Trades over Time: Dollar Volume versus Number of Trades

This figure plots the quarterly price impact of imbalance based on dollar volume (*DVOLIMB*) versus that of imbalance based on number of trades (*NIMB*) over the period 1993-2013. I follow Lee and Ready (LR, 1991) algorithm to sign trades. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped from the sample if the number of observations is less than 50. More details about data sample are provided in the Data section in Chapter 1. To estimate price impact I run the following regressions for each stock using daily observations over a quarter:

$$RETX_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \sum_{T=t,t-1} \gamma_{iT}\, DVOLIMB_{it} + e_{it}$$

$$RETX_{it} = \propto'_{it} + \beta'_{it}\, RETX_{it-1} + \sum_{T=t,t-1} \gamma'_{iT}\, NIMB_{it} + e'_{it}$$

Where $RETX_{it}$ is stock's *i* return at day *t*. I sum contemporaneous and lagged impact of imbalances in each model for each stock, and plot the cross-sectional averages of each series in each quarter. The two vertical lines refer to the period from the beginning of the fourth quarter of 2000 to the end of the second quarter of 2003.

Figure 2.4 – Stealth Trading

Figures 7A and 7B plot incremental price impact of imbalances in each bin j=1, …, 5. I estimate the following equation for each stock-quarter using daily observations:

$$RETX_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \sum_{T=t,t-1} \gamma_{iT}\, DVOLIMB_{it} + \sum_{j=1}^{5} \gamma_{ijt}\, DVOLIMBj_{it} + \sum_{j=1}^{5} \gamma_{ijt-1} DVOLIMBj_{it-1} + e_{it}$$

Where $RETX_{it}$ is stock's *i* return at day *t*, *DVOLIMB* is aggregate dollar volume imbalance, and *DVOLIMBj* is dollar volume imbalance in bin *j*= 1 (smallest), …, 5 (largest). The sum of cross-sectional averages of γ_{ijt} and γ_{ijt-1} are plotted in Figure 7A. I also calculate the difference between proportion of price impact $(\gamma_{ijt} + \gamma_{ijt-1})/\sum_j(\gamma_{ijt} + \gamma_{ijt-1})$ and proportion of volume in each bin for each stock-quarter and average the difference across stocks each quarter. Average differences are plotted in Figure 7B. I follow Barber et al.'s (2009) definition of trade size bins and Lee and Ready's (LR, 1991) algorithm to sign trades. The two vertical lines refer to the period from the beginning of the fourth quarter of 2000 to the end of the second quarter of 2003.My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped from the sample if the number of observations is less than 50. More details about data sample are provided in the Data section in Chapter 1.

Figure 2.4A – Incremental Price Impact of Imbalances

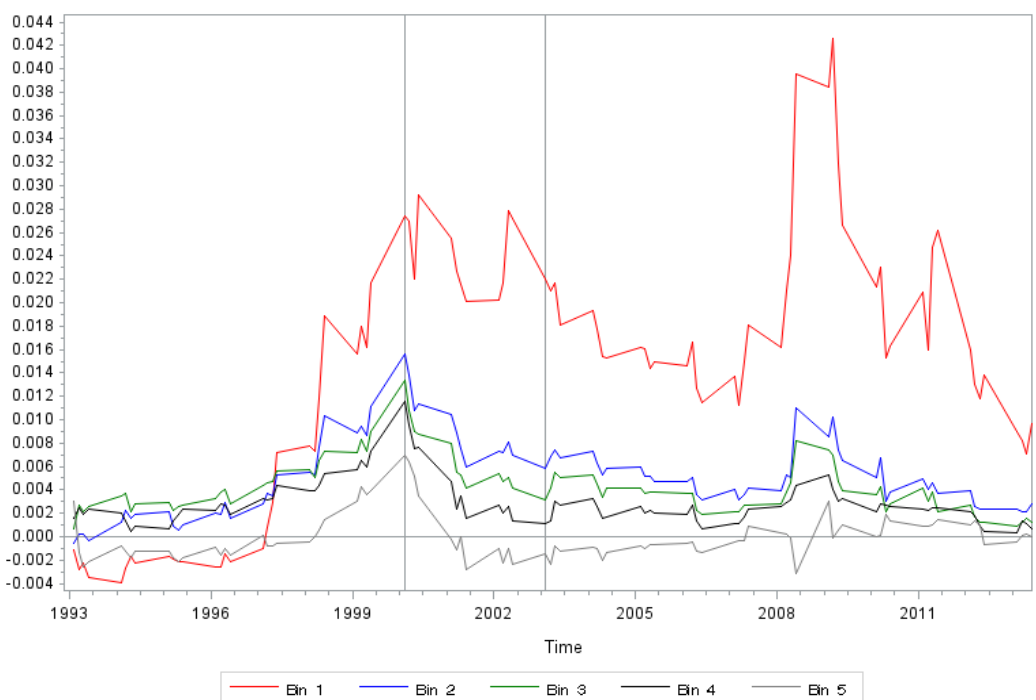


Figure 2.4B – Average Difference between Proportions of Price Impact and Volume

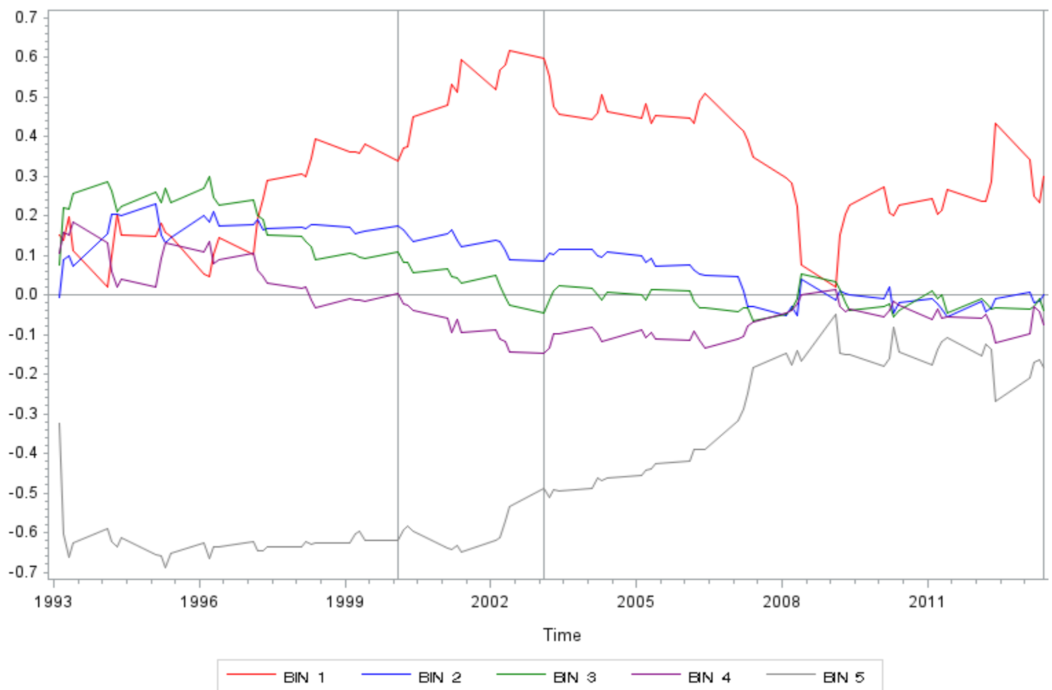


Table 2.1 - Determinants of Order Imbalances

This table presents statistics related to the coefficient estimates of the following regression equations:

$$DVOLIMB_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \beta'_{it}\, MKTRT_{t-1} + \gamma_{it}\, DVOLIMB_{it-1} + e_{it}$$

$$DVOLIMBj_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \beta'_{it}\, MKTRT_{t-1} + \sum_{j=1}^{5} \gamma_{ijt-1} DVOLIMBj_{it-1} + e_{it}$$

Results for the first (second) equation are presented in Panel A (B). First, a time-series regression is estimated for each stock in each sample. Second, cross-sectional average coefficients and t-stat of average are presented in the table. Standard errors are corrected for cross-correlations using the assumptions in Chordia and Subrahmanyam (2004), where residual correlation is proxied by the average correlation calculated from groups of 100 stocks separated based on their PERMNOs. Regressions are estimated for the early (1993-2001) and recent (2002-2013) samples. *RETX* is CRSP's daily return-excluding dividends, *MKTRT* is value-weighted average return of stocks included in my sample, and $DVOLIMBj_{it}$ is calculated as follows: $\frac{DVOLBUYj_{it} - DVOLSELLj_{it}}{DVOLBUYj_{it} + DVOLSELLj_{it}}$, where $DVOLBUYj_{it}$ ($DVOLSELLj_{it}$) is the aggregate dollar value of buy (sell)-initiated trades in size bin *j* of stock *i* at day *t,* and buy and sell trades are classified using Lee and Ready's (LR, 1991) algorithm. I follow Barber et al.'s (2009) definition of trade size bins *j*= 1 (smallest), …, 5 (largest). My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped if the number of observations is less than 50. More details about data sample are provided in the Data section in Chapter 1.

Panel A

Dependent: DVOLIMB	***Early***		***Recent***	
Variable	***Mean***	***t-stat***	***Mean***	***t-stat***
Intercept	-0.10661	-10.24	-0.04589	-9.05
$RETX_{it-1}$	-0.10811	-2.42	0.110008	0.39
$MKTRT_{t-1}$	0.81614	3.59	0.498359	3.62
$DVOLIMB_{it-1}$	0.208215	28.77	0.151356	9.78
R^2	0.1305		0.0306	

Panel B

Dependent: DVOLIMB1	***Early***		***Recent***	
Variable	***Mean***	***t-stat***	***Mean***	***t-stat***
Intercept	-0.1004	-6.11	0.003275	0.02
$RETX_{it-1}$	-0.14208	-0.22	-6.58581	-0.18
$MKTRT_{t-1}$	0.622053	0.67	0.426506	2.71
$DVOLIMB1_{it-1}$	0.190852	7.45	0.168337	1.47
$DVOLIMB2_{it-1}$	0.052052	1.80	-0.01054	-0.08
$DVOLIMB3_{it-1}$	0.02745	2.63	3.257199	0.21
$DVOLIMB4_{it-1}$	0.032038	0.42	0.119737	0.35
$DVOLIMB5_{it-1}$	-0.02866	-0.21	0.010737	0.07
R^2	0.1101		0.0348	

Dependent: DVOLIMB2	Early		Recent	
Variable	***Mean***	***t-stat***	***Mean***	***t-stat***
Intercept	-0.04562	-3.83	-0.01614	-1.98
$RETX_{it-1}$	-0.51943	-0.40	-0.38342	-0.20
$MKTRT_{t-1}$	0.393536	0.31	0.275847	1.82
$DVOLIMB1_{it-1}$	0.10192	10.25	0.07977	4.43
$DVOLIMB2_{it-1}$	0.082532	4.02	0.058477	1.59
$DVOLIMB3_{it-1}$	0.051191	2.98	-1.59232	-0.21
$DVOLIMB4_{it-1}$	0.017214	0.42	0.014683	1.27
$DVOLIMB5_{it-1}$	-0.03798	-0.18	-0.00244	-0.05
R^2	0.0727		0.0094	

Dependent: DVOLIMB3	Early		Recent	
Variable	***Mean***	***t-stat***	***Mean***	***t-stat***
Intercept	-0.02962	-3.37	0.125798	0.20
$RETX_{it-1}$	-0.06568	-0.08	-0.64594	-0.19
$MKTRT_{t-1}$	0.794624	0.96	0.114439	0.67
$DVOLIMB1_{it-1}$	0.059285	8.56	0.185499	0.29
$DVOLIMB2_{it-1}$	0.047099	3.37	-0.10112	-0.16
$DVOLIMB3_{it-1}$	0.057407	4.84	-0.08642	-0.14
$DVOLIMB4_{it-1}$	0.034265	7.55	0.291527	0.23
$DVOLIMB5_{it-1}$	0.017082	0.95	-0.15514	-0.24
R^2	0.0442		0.0068	

Dependent: DVOLIMB4	Early		Recent	
Variable	***Mean***	***t-stat***	***Mean***	***t-stat***
Intercept	-0.01673	-3.37	0.030675	0.17
$RETX_{it-1}$	-0.26337	-0.64	1.203654	0.18
$MKTRT_{t-1}$	0.415176	0.76	0.037709	0.23
$DVOLIMB1_{it-1}$	0.030891	5.17	0.079555	0.45
$DVOLIMB2_{it-1}$	0.029125	3.42	-0.01394	-0.08
$DVOLIMB3_{it-1}$	0.037092	3.69	-0.88572	-0.21
$DVOLIMB4_{it-1}$	0.0488	6.92	0.105067	0.30
$DVOLIMB5_{it-1}$	0.078103	0.32	-0.0259	-0.14
R^2	0.0287		0.005	

Dependent: DVOLIMB5	Early		Recent	
Variable	***Mean***	***t-stat***	***Mean***	***t-stat***
Intercept	-0.00962	-0.42	0.026002	0.22
$RETX_{it-1}$	0.160501	0.96	-6.74754	-0.21
$MKTRT_{t-1}$	0.185207	1.44	0.024731	0.16
$DVOLIMB1_{it-1}$	0.000468	0.02	0.01069	0.20
$DVOLIMB2_{it-1}$	0.006865	1.91	0.017452	0.52
$DVOLIMB3_{it-1}$	0.01029	2.11	0.09188	0.32
$DVOLIMB4_{it-1}$	0.025297	0.86	0.038105	0.42
$DVOLIMB5_{it-1}$	0.051863	14.22	0.046283	0.52
R^2	0.0155		0.0032	

Table 2.2 - Price Impact of Order Imbalances

This table presents statistics about coefficient estimates of the following regression equations:

$$RETX_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \sum_{T=t,t-1} \gamma_{iT}\, DVOLIMB_{it} + e_{it}$$

$$RETX_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \sum_{j=1}^{5} \gamma_{ijt}\, DVOLIMBj_{it} + \sum_{j=1}^{5} \gamma_{ijt-1} DVOLIMBj_{it-1} + e_{it}$$

Results for the first (second) equation are presented in Panel A (B). First, a time-series regression is estimated for each stock in the early and recent samples. Second, cross-sectional average coefficients and t-stat of average are presented in the table. Standard errors are corrected for cross-correlations using the assumptions in Chordia and Subrahmanyam (2004), where residual correlation is proxied by the average correlation calculated from groups of 100 stocks separated based on their PERMNOs. Regressions are estimated for the early (1993-2001) and recent (2002-2013) samples. *RETX* is CRSP's daily return-excluding dividends, and $DVOLIMB_{it}$ is calculated as follows: $\frac{DVOLBUY_{it} - DVOLSELL_{it}}{DVOLBUY_{it} + DVOLSELL_{it}}$, where $DVOLBUY_{it}$ ($DVOLSELL_{it}$) is the dollar value of buy (sell)-initiated trades of stock *i* at day *t,* and buy and sell trades are classified using Lee and Ready's (LR, 1991) algorithm. $DVOLIMBj_{it}$ is calculated analogously within each bin *j* where *j* = 1 (smallest), …, 5 (largest) are according to the definition provided by Barber et al. (2009). My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped if the number of observations is less than 50. More details about data sample are provided in the Data section in Chapter 1.

Panel A

	Early		*Recent*	
Variable	*Mean*	*t-stat*	*Mean*	*t-stat*
Intercept	0.0036	8.48	0.0025	7.95
$RETX_{it-1}$	-0.0838	-14.92	-0.0582	-11.72
$DVOLIMB_{it}$	0.0296	20.54	0.0384	23.31
$DVOLIMB_{it-1}$	-0.0041	-8.48	-0.0065	-11.19
R^2	0.0538		0.0576	

Panel B

	Early		*Recent*	
Variable	*Mean*	*t-stat*	*Mean*	*t-stat*
Intercept	0.0035	7.84	0.0023	7.17
$RETX_{it-1}$	-0.0852	-14.41	-0.0555	-10.67
$DVOLIMB1_{it}$	0.0210	13.42	0.0425	23.74
$DVOLIMB1_{it-1}$	-0.0040	-6.85	-0.0080	-12.05
$DVOLIMB2_{it}$	0.0087	12.33	0.0081	9.71
$DVOLIMB2_{it-1}$	-0.0003	-0.69	-0.0004	-0.67
$DVOLIMB3_{it}$	0.0078	11.91	0.0062	6.52
$DVOLIMB3_{it-1}$	-0.0004	-0.77	-0.0006	-0.96
$DVOLIMB4_{it}$	0.0066	8.83	0.0046	3.90
$DVOLIMB4_{it-1}$	-0.0006	-0.93	-0.0010	-1.30
$DVOLIMB5_{it}$	0.0039	4.37	0.0013	1.04
$DVOLIMB5_{it-1}$	-0.0018	-2.31	-0.0013	-1.66
R^2	0.0528		0.0711	

Table 2.3 - Price Impact of Order Imbalances – R^2

This table presents average R-squared values from regressing return-excluding-dividend ($RETX_{it}$) on lagged firm's return ($RETX_{it-1}$) in addition to different combinations of contemporaneous and lagged imbalances in different trade size bins j, where j = 1 (smallest), …, 5 (largest) is according to the definition of Barber et al. (2009). Dollar volume imbalance ($DVOLIMBj_{it}$) is calculated as follows: $\frac{DVOLBUYj_{it} - DVOLSELLj_{it}}{DVOLBUYj_{it} + DVOLSELLj_{it}}$, where $DVOLBUYj_{it}$ ($DVOLSELLj_{it}$) is the aggregate dollar value of buy (sell)-initiated trades in size bin j of stock i at day t, and buy and sell trades are classified using Lee and Ready's (1991) algorithm. Regressions are estimated for each stock then cross-sectional average R-squared values are presented. In the table below under "Variables" I list independent variables in the regression equation. For instance, when Bin 1 is included, this means that $DVOLIM1_{it}$ and $DVOLIMB1_{it-1}$ are included, and so on. When Bin 3 and Bin 4 are included, this means that $DVOLIM3_{it}$, $DVOLIMB3_{it-1}$, $DVOLIM4_{it}$ and $DVOLIMB4_{it-1}$ are included in the regression, and so on. Regressions are estimated for the early (1993-2001) and recent (2002-2013) samples. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped if the number of observations is less than 50. More details about data sample are provided in the Data section in Chapter 1.

Variables	Early	Recent
$RETX_{it-1}$	0.0131	0.0041
$RETX_{it-1}$+ Bin 1	0.031	0.0613
$RETX_{it-1}$+ Bin 2	0.028	0.0153
$RETX_{it-1}$+ Bin 3	0.033	0.0101
$RETX_{it-1}$+ Bin 4	0.0203	0.0068
$RETX_{it-1}$+ Bin 5	0.0154	0.0043
$RETX_{it-1}$+ Bin 3 + Bin 4	0.0412	0.012
$RETX_{it-1}$+ Bin 2 + Bin 3	0.0431	0.0191
$RETX_{it-1}$+ Bin 2 + Bin 3 + Bin 4	0.0485	0.0204
$RETX_{it-1}$+ Bin 1 + Bin 5	0.0394	0.0614
$RETX_{it-1}$+ Bin 1 + Bin 2 + Bin 3 + Bin 4	0.0504	0.0674
$RETX_{it-1}$+ Bin 2 + Bin 3 + Bin 4 + Bin 5	0.0366	0.0205
$RETX_{it-1}$+ Bin 1+ Bin 2 + Bin 3 + Bin 4 + Bin 5	0.0539	0.0675

Chapter 3 – On the Adaptability of Informed Traders to Changing Market Conditions

Introduction

Evidence presented in the first two chapters of this thesis shows that the size distribution of trades shifted significantly toward smaller sizes over the sample period, and that price impact transitioned in the same direction. The close association between the shift of the distribution and transition of permanent price impact, as demonstrated by the results presented so far, may indicate that informed traders are directly involved in those change patterns. The purpose of this chapter is to analyze the behavior of informed traders in more detail during my sample period. A number of questions are addressed to that effect.

First, which type of trades (i.e. medium or large trades) is primarily responsible for the migration of volume towards the small bin? Earlier studies such as that of Barclay and Warner showed that most of the information-based trades were of medium size. If informed traders were actively involved in the migration of volume towards small size, I expect to find that the increase in small trading volume is associated with a decrease in medium trading volume in particular. My results for this test point in that direction.

I then proceed to examine in more detail the activity of informed traders. If informed traders are now strategically trading in small sizes, their activity implies that trading outcomes differ for stocks with different probabilities of information. To validate this conjecture, I examine trading activity for stocks conditional on their Probability of Information-based Trading (PIN), a measure closely related to imbalance developed by Easley et al. (2002). My results show that the relationship between dollar volume and PIN is generally negative, even after controlling for liquidity. In terms of the direction of this relationship within each trade, I find that in recent (early) years the relation between the proportion of medium trades and PIN is negative (positive). For small trades, the relation between proportion of volume and PIN is positive, but the relation becomes more pronounced in recent years, consistent with informed trading migrating to small trades in that period.

My findings also have implications for the theoretical and empirical microstructure literature. O'Hara (2015) calls for a new microstructure research agenda because traditional theoretical and empirical methods might no longer be appropriate in the light of recent changes in markets. My study can be considered a step in this direction. It draws attention to the increasing importance of small trades. Results also point to a diminishing role that the size of trade can play in conveying a signal about the nature of trades. It seems that the positive correlation between trade size and its informational content (e.g. Easley and O'Hara, 1987) has vanished or even reversed.

Finally, I examine whether the documented patterns in the distribution of trades and the transition of price impact are due to new trading activity, a changing strategy on the part of existing traders, or both factors. I examine cases in which trading volume increases or decreases significantly. Such extreme volume change cases might represent a change in trading participants. I find that in those cases the effect is reflected mainly in large trades. I conclude that the patterns documented are primarily due to a shift in the strategies employed by existing traders. In addition, when the proportions of volume in each bin change significantly, price impact transitions across bins in a similar fashion and in the same month when the shift in volume takes place. This finding confirms that informed traders make decisions dynamically according to changes in market conditions.

The rest of the chapter is organized as follows. Section 3.1 examines the interaction of trading volume between different bins. Section 3.2 elaborates on the role of information-based trading in driving my results. Section 3.3 discusses the role of significant changes in volume and the optimality of trading strategies. Section 3.4 offers conclusions based on these insights.

3.1 Tracing Information-based Trading

In this section, I attempt to trace informed trading activity across bins. Figures presented so far indicate that there might be some interaction in activities across bins. Of course, as the proportion of volume increases in Bin 1, proportions of volume decrease in other bins. I attempt first to identify which bins experience most of the loss in volume to Bin 1. Second, I examine changes in the price impact coefficients of Bins 2-5 as price impact increases for Bin 1. Answering these two questions helps us to follow the migration of trading volume in general and informed volume in particular.

To address the first question, I calculate correlations in each quarter between the change in the proportion of dollar volume attributed to Bin 1 and those related to Bins 2-5. In each quarter I calculate, at the stock level, the difference (with respect to the preceding quarter) in proportion of volume in each bin; correlation coefficients are then obtained for each quarter. In quarters when the proportions of volume attributed to Bin 1 increase, I expect the proportions of medium-sized bins in particular to decrease, if migration involves informed traders and informed trading is concentrated in medium bins in the early years. Therefore, I expect to observe the most negative correlations between Bin 1 and the medium-sized bins. To address the second relation, in each quarter I estimate equation 4 for each stock, find in each bin the quarterly change in price impact for each stock, and calculate correlations between changes in the price impact of Bin 1 and those of Bins 2-5. In quarters when the price impact of Bin 1 increases, I expect this increase to result from the migration of informed traders from medium-sized bins, causing the price impact of those bins to decrease. Therefore, I expect to see the lowest correlations between Bin1 and medium-sized bins.

The two sets of correlations are plotted in Figure 3.1. The figure titled "BIN 2" plots correlations of changes in proportion of *DVOL* and price impact between Bin 1 and Bin 2. Similarly, the figures titled "BIN 3", "BIN 4", and "BIN 5" plot correlations between Bin 1, and Bins 3, 4, and 5, respectively. Unsurprisingly, correlations of volume proportion are all negative. *Ceteris paribus*, if volume increases in one bin, then proportion will increase in that bin and decrease in all other bins. However, the most negative correlations are to be seen between order imbalance in Bin 1 and Bin 2 (in recent years), followed by Bin 3, and Bin 4. In addition, the decrease in correlation is sharper for medium-sized bins in the highlighted area.

Results presented in Chapter 1, such as Figure 1.2, show that proportions of volume shifted over the sample period. Specifically, Figure 1.2 shows that the proportion of volume contained in Bin 1 increased significantly, the proportion in Bin 5 decreased significantly, and the proportions of volume in other bins stayed relatively stable. Results in this section improve our understanding of the nature of this shift by showing that, in recent years, volume migrated to Bin 1 primarily from medium-sized bins, and that the Bin 5 to Bin 1 shift documented in Chapter 1 represents a gradual reduction in trade sizes. Price impact correlations are noisy and a conclusion is not obvious, though correlations with Bin 2 seem to drift below zero most often.

3.2 Probability of Information-based Trading

If informed traders are strategically trading in small sizes, we may presume that trading style differs for stocks with a different probability of information-based trading. To validate this conjecture, I examine trading activity for stocks conditional on their Probability of information-based trading (PIN), a measure developed by Easley et al. (1997) that is closely related to trade imbalance. The idea is that rates of informed and uninformed trades can be inferred from the imbalance between buy and sell trades. Specifically, PIN = $\alpha\mu/(\alpha\mu + 2\varepsilon)$, where α is the probability of an information event, μ is the arrival rate of informed traders, and ε is the arrival rate of uninformed traders. The numerator represents the expected number of informed trades and the denominator is the expected number of informed and uninformed trades. Arrival rates μ and ε are estimated from the imbalance between buy and sell trades, and α is estimated from the proportion of days with abnormal trading activity.

I obtain quarterly estimates of PIN[20] and rank stocks each quarter into quintiles based on PIN. In Table 3.2, I report PIN averages, dollar volumes, and the proportion of average volume in each bin and for six three-year periods from 1993 to 2010 (PIN estimates are available until 2010). A number of observations can be made based on Table 3.2. First, dollar volume decreases in PIN in all sub-periods. Stocks with higher probability of informed trading tend to be less liquid; those are smaller stocks (firm size unreported) with information less readily available for investors and therefore the ratio of private-to-public information is higher.

Second, regarding the proportions of average volume in individual bins: if informed traders trade strategically in certain sizes and avoid others, it is expected that the proportion of volume

[20] I thank Stephen Brown for making PIN estimates available on his website. PIN is estimated using an extended version of the model. This is available at: http://scholar.rhsmith.umd.edu/sbrown/pin-data.

increases (decreases) in PIN in bins that traders prefer (avoid). The reason for this conjecture is that liquidity traders are not expected to trade strategically; their trades exhibit less trade size bias. In addition, their trading volume is relatively stable and the distribution of their trades is not expected to change significantly over short horizons. On the other hand, the other main type of trading—informed trading—intensifies around high information periods (identified by high PIN quarters), and informed traders select their trade sizes strategically. The result of this combination is that high PIN periods are associated with an increase in the proportion of volume contained in the bins that informed traders prefer. It is not necessarily the case that volume itself increases in those bins, because high-information risk periods might experience reduced activity on the part of uninformed traders.

My analysis so far suggests that informed traders in the early years of the study traded more in medium-sized trades and less in large or small trades, but that, in recent years, informed trading shifted to small bins, whereas less information-based trading occurred in medium and large bins. Therefore, we should find that (1) in the large bin, the proportion of volume decreases in PIN both in early and in recent years; (2) in medium-sized bins, the proportion increases in PIN in the early years and decreases in recent years; and (3) in the small bin, it decreases in PIN in the early years and increases in PIN in recent years.

In Table 3.1 I report the average trading volume in each PIN quintile during six sub-periods. Stocks are sorted into PIN quintile in each quarter; averages are first calculated on the stock-level in each quarter, and then averages across stock-quarters are presented. The results, shown in Table 3.2, are largely consistent with my predictions. The proportions of volume in Bin 5 decrease in PIN quintiles in all sub-periods except in 2008-2010, when they decrease and then increase again. The proportions in medium-sized Bins 2, 3, and 4 increase in PIN quintiles in the early years and decrease in later years. Note that the shift in pattern takes place in 1996-1998 for Bin 4, in 2002-2004 for Bin 3, and in 2005-2007 for Bin 2. This is in line with the hypothesis that informed traders gradually move from larger to smaller trade sizes. The results for Bin 1 are slightly different than expected; the proportions of volume increase in PIN in all sub-periods, though the extent of increase is larger in recent years. This unanticipated pattern for Bin 1 probably occurs because Bin 1 is not the last in order when it comes to informed trading in early years, as can be seen in Figure 2.2, where its price impact is just under that of medium bins.

3.3 On the Optimality of Trading Strategies

As mentioned above, evidence presented so far points to a shift in the trade size distribution of volume towards smaller sizes accompanied by a similar transition in price impact, as well as a strong association between the two changes. Figure 3.2 presents five graphs (one for each trade size bin) in which proportions of volume and price impact coefficients (the sum of contemporaneous and lagged coefficients of order imbalance in Models 3 and 4) are presented together for each year of the sample. For Bin 1, in particular, the proportion of volume and the strength of price impact exhibit a positive association throughout the years.

As mentioned above, using Easley and O'Hara's (1987) theory leads to the prediction that market is in a pooling equilibrium currently and that this implies that informed traders followed the shift in the distribution of trades. Therefore, the association between the shift in proportional volume and the transition in price impact is anticipated. However, the *extent* of price impact transition is very large; in recent years of the study, the permanent price impact exerted by small trades of Bin 1 and Bin 2 increased to 0.035 and 0.0077, respectively (Table 2.2). On the other hand, the permanent price impact of large trades in Bin 4 decreased to 0.0036, whereas the price impact of Bin 5 trades became almost nonexistent.

The extent of price impact transition across bins might leave the reader wondering whether traders have reacted optimally to market changes. That is, trading activity in the large trade category declined even though a smaller price impact was associated with it in recent years. This observation is particularly puzzling for informed traders who are expected to be more sophisticated and to trade optimally with respect to minimizing price impact. We do observe a slight reverse migration of volume and reverse transition of price impact in the most recent years of the sample, but efficient trading behavior dictates more immediate and matching adjustments. Nevertheless, it is not possible to conclude based on this evidence alone that traders act in an inefficient manner. One alternative explanation for this observation is the potential presence of restrictions on trading in large or even medium sizes; this could explain why traders do not trade in large sizes or take advantage of reduced price impact. Such potential restrictions may include lower depths and higher transparency. The presence of such restrictions means that if large trades were to be placed, the price impact would jump materially.

To shed more light on the optimality of trading behavior I examine cases in which trading volume experienced a significant change—stock-months with a minimum of either a 50% increase or a 50% decrease in dollar volume. Instances of significant volume increase may indicate the presence of new participants in the market or large changes in demands of existing investors. Put differently, in the absence of significant volume changes, it seems unlikely that a change has occurred in the market participation pool. An additional purpose of this volume change analysis is to discern whether changes in the size distribution of trades and price impact pattern are due to new market participants or to shifting strategies on the part of existing traders.

In this section, I aggregate data to the monthly level to improve my ability to detect significant changes in trading volume. Examining changes in volume at the daily level is subject to noise. Also, in contrast to the analysis above in which I consider five trade size classifications, my analysis in this section is restricted to three such classifications: small (merger of Bins 1 and 2), medium (Bin 3), and large (merger of Bins 4 and 5). This consolidated bin design is employed to restrict the focus to the share of small versus non-small trades, rather than to inter-category changes.

Table 3.2 reports the number of observations in which a stock's volume has increased/decreased by 50%; those numbers are 294,091 and 168,541 for cases of increases and decreases, respectively. These numbers of observations constitute 23% and 13% of the sample, respectively. I report those percentages in early and recent years separately; the early years range from 1993 to 2001, and the recent years comprise 2002 to 2013. The frequencies of both volume increase and decrease are lower in recent years; the percentage of observations that are classified as increases (decreases) is 25% (16%) in early years versus 20% (9%) in recent years. In addition, I split stocks into terciles based on the market capitalization as of the last trading day in the previous year. I report those percentages for the three market capitalization portfolios, in the early and recent years separately. The percentages fall monotonically in market capitalization; that is, trading activity is more stable for larger stocks.

In Table 3.3 I report the cross-sectional averages of monthly return, and for each trade size category I indicate the dollar volume, the change in proportions of dollar volume, and the change in proportions of order imbalance. The proportion of order imbalance is the proportion of imbalance in the direction of return (which may be considered as an indicator of the proportion of information-based volume) for each stock-month in each trade size bin $k \in \boldsymbol{K}$, where $\boldsymbol{K}$ = {small, medium, large}. It is calculated as follows:

$$(DVOL\ Buy_k - DVOL\ Sell_k) * D_k / \sum_{k \in \boldsymbol{K}} (DVOL\ Buy_k - DVOL\ Sell_k) * D_k$$

Where D_k is a dummy variable that equals 1 if the direction of order imbalance is in the direction of return (i.e. positive imbalance when return is positive and vice versa) and 0 otherwise. Panel A (B) of Table 3.3 reports those statistics for volume increase (decrease) cases, and in each panel the statistics are reported for the month of change (Month 0), the preceding month (Month -1) and the following month (Month 1), for the whole sample period and for early and recent years separately. We may note first that there is a positive relation between volume change and return and that this relation is stronger in early years. Months witnessing significant volume increase have high returns of 8% on average, and those witnessing significant volume decrease have low returns of -4.8% on average. This evidence about the relationship between volume and return is similar to the findings reported by Gervais et al. (2001).

Examining dollar volume patterns shows that those significant volume increase and decrease cases are reflected in all trade sizes, but are concentrated in the large trade size category. In fact, about 50% of the change in volume is reflected in the large bin, followed by the small bin and then the medium bin. This asymmetric change across bins creates a change in the distribution of trades. In the case of significant volume increase (decrease), the proportion of volume in large trades increases (decreases) by about 10% (14%) in Month 0, while the proportion of volume in small trades decreases (increases) by about 10% (13%), and the change in the proportion of medium size trades is insignificant.

If information-based traders are heavily involved in significant volume change cases, and if informed traders seek to minimize trading costs, then volume increases should be reflected primarily in large trades, because large trades were associated with the smallest price impact in the later years of the study. Similarly, volume decrease cases should be reflected in small trades, since this is where informed traders concentrate their price-moving trades. Based on these assumptions, the proportion of large trades should increase in the case of volume increase and the proportion of small trades should decrease in the case of volume decrease. The findings about the change in proportions of volume are consistent with this prediction regarding volume increase, but not in the case of volume decrease.

It is possible that the predicted patterns are not observed because volume increases are dominated by urgent demands for trading and that volume decreases coincide with thinning trading. In this case, examining the changes in the proportion of order imbalance and in the price impact of trades in different bins might exhibit the above-mentioned predictions, even if information-based traders are involved in those changes. However, results in the table show that the change in order imbalance is also very closely related to that in the size distribution of trades. This result indicates that the evidence presented in this thesis about the leftward shift in the distribution of trades and the accompanying transition in price impact are due primarily to the redistribution of existing traders. Finally, the bottom rows of Table 3.3 report the price impact coefficients in each trade size category, along with t-stat values[21]. In the case of significant volume increase (decrease), the price impact associated with small trades also increases (decreases) during the month of change. The effect of significant volume change on the price impact of medium and large trades moves in the same direction, but is of a much smaller magnitude than that for small trades. Since the price impacts of all trades move in the same direction in cases of both increase and decrease, a conclusion about any transition in price impact

[21] Price impact coefficients are estimated using a panel regression whereby monthly returns are regressed on lagged monthly returns and contemporaneous order imbalance in each of the three trade size categories. Coefficients of order imbalance and t-statistics are reported in the table. Standard errors are clustered by stock and month.

is not obvious. In other words, it is unclear whether the change in price impact relative to the change in volume is positive or negative.

In summary, the tests performed in this section suggest that the volume shift and the change in price impact pattern are not due to new market participants, because those patterns are not found when examining cases with significant changes in trading volume. However, none of the tests is conclusive about whether traders seek optimal trading strategies with respect to minimizing trading costs.

3.4 Conclusions

After characterizing the changes in trading activity (Chapter 1) and price impact pattern (Chapter 2), I attempt in Chapter 3 to analyze in more detail the change in informed trading activity over time. I find that while all trade size bins lost ground to the smallest bin, the interchange of volume was strongest between the small and medium-sized bins, as evident in the strong negative correlations between the change in the small trade size bin and that of medium bins.

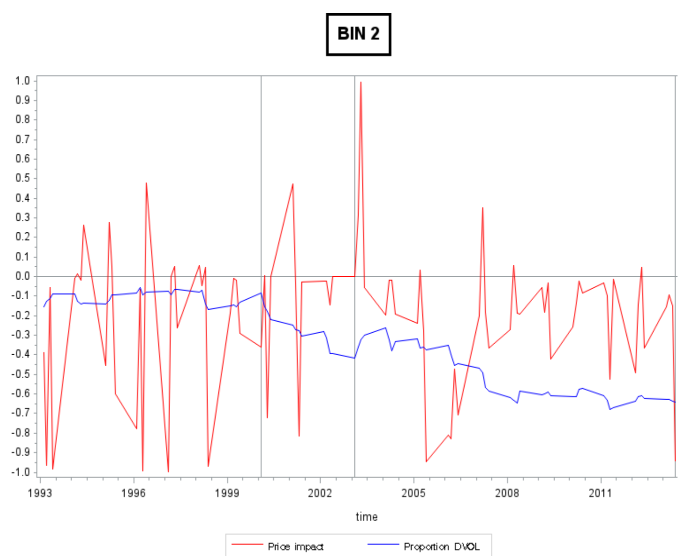
I also test whether a temporary increase in information-based trading shifts the distribution of trades towards small sizes. Probability of information-based trading (PIN) is a measure for the risk of trading against an informed trader and it is, by construction, closely related to order imbalance. I classify stocks according to their PIN values each quarter, and I find that stocks with high PIN values tend to have higher fraction of small trade size volume. This result holds even after I control for average stock price, which could influence the typical trade size of a stock.

Figure 3.1– Correlations of Changes in Price Impact and Proportion of Volume between Bin 1 and Other Bins

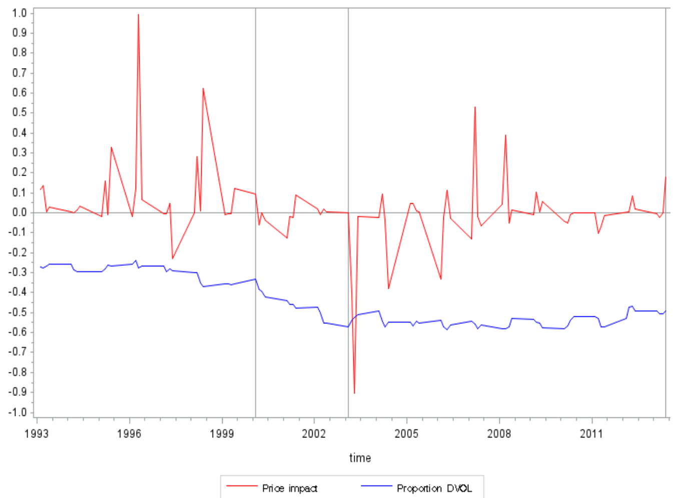
This figure plots quarterly correlations of quarterly changes in the price impact of trades and proportion of volume between Bin 1 and each of the other bins. Quarterly price impact is estimated from the following equation:

$$RETX_{it} = \propto_{it} + \beta_{it}\, RETX_{it-1} + \sum_{j=1}^{5} \gamma_{ijt}\, DVOLIMBj_{it} + \sum_{j=1}^{5} \gamma_{ijt-1} DVOLIMBj_{it-1} + e_{it}$$

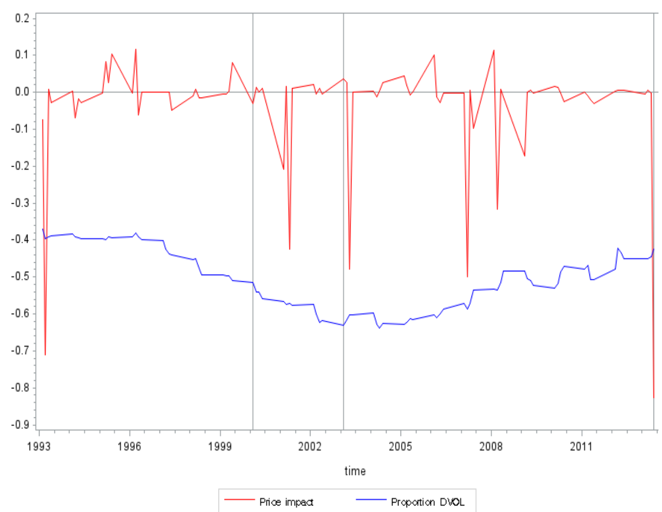
Where $RETX_{it}$ is stock's i return at day t. Price impact is the sum of contemporaneous and lagged imbalance effects (γ_{ijt} and γ_{ijt-1}) in each bin j = 1 (smallest), …, 5 (largest), for each stock. I also calculate average quarterly proportions of dollar volume (*DVOL*) in each bin for each stock. I calculate changes in price impact and proportion of volume by finding the quarterly differences in those values. Next, in each quarter correlations across stocks are calculated between Bin 1 and each of the other bins for both change in price impacts and change in proportions of volume and plotted in the graph below. I follow Barber et al.'s (2009) definition of trade size bins and Lee and Ready's (LR, 1991) algorithm to sign trades. The two vertical lines refer to the period from the beginning of the fourth quarter of 2000 to the end of the second quarter of 2003. My sample covers the period 1993-2013 and is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped from sample if the number of observations is less than 50. More details about data sample are provided in the Data section of Chapter 1.



BIN 3



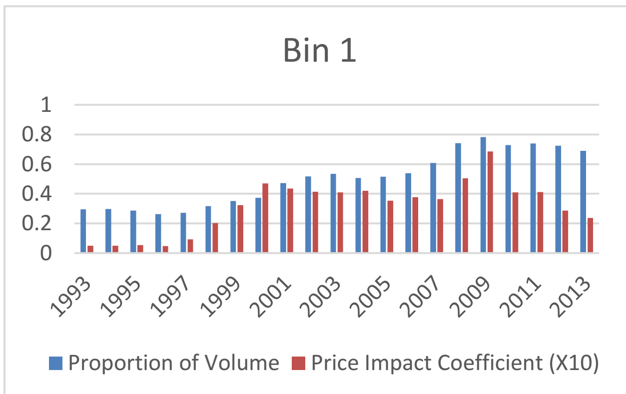
BIN 4

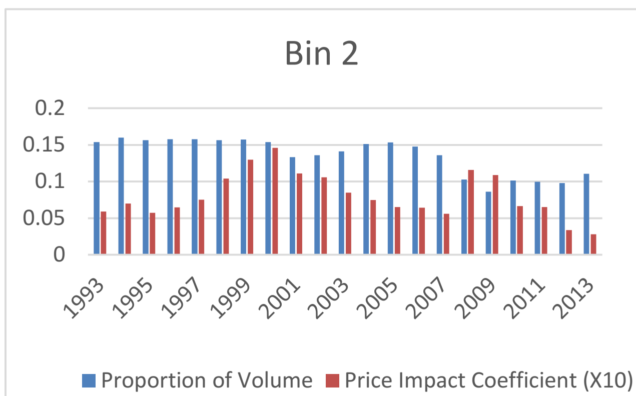


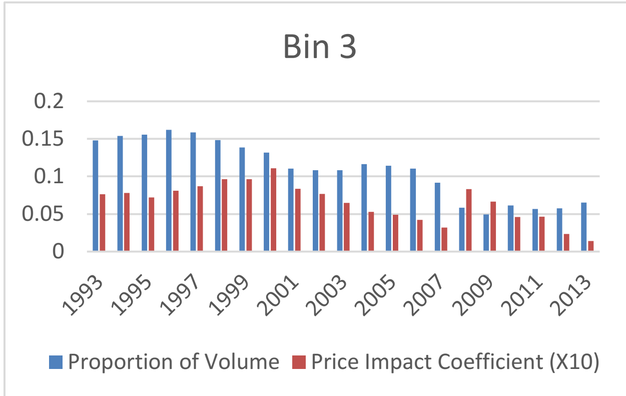
BIN 5

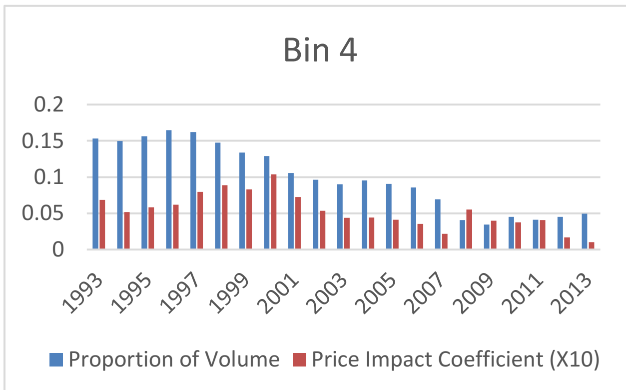
Figure 3.2 – Proportion of Volume versus Price Impact Coefficient in Each Bin

This figure plots the proportion of volume and permanent price impact for each trade size bin in each year during the sample period. Price impact is the sum of contemporaneous and lagged imbalance effects (γ_{ijt} and γ_{ijt-1}) in Model 4, in each bin j = 1 (smallest), …, 5 (largest). I follow Barber et al.'s (2009) definition of trade size bins and Lee and Ready's (LR, 1991) algorithm to sign trades. My sample covers the period 1993-2013 and is the intersection between TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped from the sample if the number of observations is fewer than 50. More details about data sample are provided in the Data section of Chapter 1.






Bin 3
0.2
0.15
0.1
0.05
0
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
Proportion of Volume
Price Impact Coefficient (X10)


Bin 4
0.2
0.15
0.1
0.05
0
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
Proportion of Volume
Price Impact Coefficient (X10)

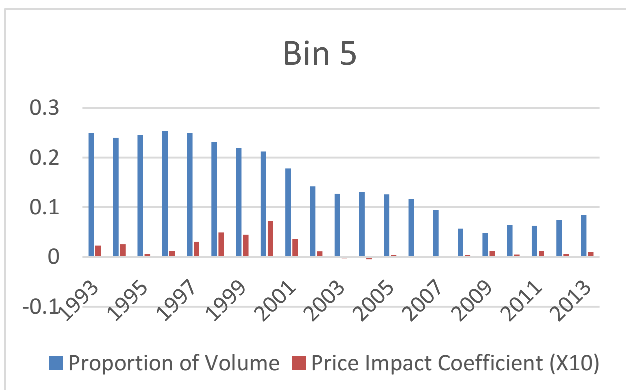

Bin 5
0.3
0.2
0.1
0
-0.1
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
Proportion of Volume
Price Impact Coefficient (X10)

Table 3.1 - Trading Volume and Probability of Informed Trading (PIN)

This table presents statistics for portfolios of stocks sorted into quintiles by quarterly Probability of Informed Trading (PIN) in six three-year subperiods. I present in each PIN portfolio the following statistics: average PIN value, average dollar volume (*DVOL*), and average proportions of volume in five trade size bins (1 is smallest and 5 is largest) as defined by Barber et al. (2009). Averages are first calculated on the stock-level in each quarter, and averages across stock-quarters are then presented. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped if the number of observations is less than 50. More details about data sample are provided in the Data section in Chapter 1.

		MEAN		***Proportion DVOL***				
Subperiod	***PIN Quintile***	***PIN***	***DVOL***	***Bin 1***	***Bin 2***	***Bin 3***	***Bin 4***	***Bin 5***
	1	0.10	451,420,523	1.25%	2.40%	4.56%	12.33%	79.46%
	2	0.18	189,761,069	2.11%	3.38%	6.20%	12.80%	75.51%
1993-1995	3	0.24	83,798,057	2.96%	4.42%	7.56%	13.80%	71.26%
	4	0.33	34,840,129	3.78%	5.03%	8.26%	14.31%	68.62%
	5	0.56	9,318,951	5.05%	5.98%	8.87%	14.12%	65.98%
	1	0.10	1,125,453,300	1.28%	2.70%	5.40%	15.09%	75.53%
	2	0.17	266,451,750	2.61%	4.12%	7.46%	14.69%	71.11%
1996-1998	3	0.23	92,176,634	4.02%	5.42%	8.73%	15.09%	66.73%
	4	0.31	33,203,727	5.49%	6.55%	9.55%	15.11%	63.30%
	5	0.52	10,203,618	6.20%	6.84%	9.89%	14.78%	62.29%
	1	0.09	2,643,223,459	3.13%	5.01%	8.47%	16.14%	67.24%
	2	0.16	430,778,581	4.87%	5.54%	8.30%	14.25%	67.04%
1999-2001	3	0.23	81,014,440	10.05%	9.02%	10.96%	14.71%	55.26%
	4	0.32	19,823,429	15.13%	11.44%	12.89%	15.12%	45.42%
	5	0.52	6,115,475	14.58%	11.23%	12.51%	14.91%	46.77%
	1	0.10	2,748,082,963	9.75%	10.07%	14.11%	18.43%	47.64%
	2	0.16	354,346,050	20.67%	13.23%	14.09%	15.76%	36.24%
2002-2004	3	0.22	88,562,621	30.87%	13.81%	12.71%	13.07%	29.54%
	4	0.31	27,246,982	35.12%	13.33%	11.88%	12.38%	27.28%
	5	0.49	7,824,359	30.83%	13.17%	12.96%	13.28%	29.76%
	1	0.08	4,353,569,427	18.45%	14.85%	15.54%	18.51%	32.64%
	2	0.13	894,369,021	29.33%	15.54%	14.54%	14.23%	26.36%
2005-2007	3	0.18	313,733,304	33.65%	14.35%	13.33%	12.48%	26.19%
	4	0.26	74,907,085	38.88%	12.99%	11.56%	11.36%	25.21%
	5	0.45	20,761,492	35.61%	13.94%	12.79%	12.71%	24.96%
	1	0.06	6,714,926,981	40.73%	17.85%	13.90%	11.78%	15.73%
	2	0.11	983,335,368	54.48%	13.74%	9.94%	8.02%	13.83%
2008-2010	3	0.17	283,883,392	55.83%	12.37%	9.23%	8.02%	14.55%
	4	0.27	73,219,147	56.83%	12.12%	8.99%	7.98%	14.08%
	5	0.46	26,672,242	47.82%	12.73%	9.64%	9.15%	20.66%

Table 3.2 - Proportions of Significant Volume Change Observations

This table provides some statistics about stock-month observations (1,300,137 observations) that experienced significant changes—that is, either a minimum 50% increase ("Increase") or 50% decrease ("Decrease")—in trading volume, compared to the previous month. The table presents percentages of the total number of observations in each case, for the entire sample period (1993-2013), the early (1993-2001) sub-sample, and the recent (2002-2013) sub-sample. In addition, percentages are presented in three market capitalization categories (Small, Mid, and Large) in the early and recent periods. Stocks are classified into market capitalization terciles based on the last trading day in the previous year. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped if the number of observations is fewer than 50. More details about data sample are provided in the Data section in Chapter 1.

				Early			Recent		
	Overall	**Early**	**Recent**	**Small Cap**	**Mid Cap**	**Large Cap**	**Small Cap**	**Mid Cap**	**Large Cap**
Increase	22.62%	24.87%	19.58%	28.58%	25.53%	20.69%	25.04%	19.69%	14.22%
Decrease	12.96%	15.66%	9.31%	20.73%	15.94%	10.56%	15.35%	8.56%	4.25%

Table 3.3 - Descriptive Statistics about Significant Volume Change Observations

This table provides statistics related to stock-month observations (1,300,137 observations) that experienced significant changes of either 50% increase ("Increase") or 50% decrease ("Decrease") in trading volume, compared to the previous month. Statistics are provided for the month of change (0), the preceding month (-1) and the following month (1), for the entire sample period (1993-2013), the early (1993-2001) sub-sample, and the recent (2002-2013) sub-sample. Statistics are provided for small (Bins 1 and 2), medium-sized (Bin 3), and large (Bins 4 and 5) trades, where bins are defined following Barber et al.'s (2009) definition. The table reports averages of return, dollar volume, change in proportion of imbalance, and price impact. Proportion of order imbalance is based on imbalance-in-the-direction-of return. Price impact coefficients are predicted by estimating a regression model similar to Model 4. Panel A (B) presents statistics for Increase (Decrease) cases. My sample is the intersection of TAQ and CRSP data sets, excluding non-ordinary stocks. In addition, stocks are dropped if the number of observations is fewer than 50. More details about data sample are provided in the Data section in Chapter 1.

		1993-2013			1993-2001			2002-2013		
Panel A – Increase		**-1**	**0**	**1**	**-1**	**0**	**1**	**-1**	**0**	**1**
Return		-0.004	0.080	-0.006	-0.012	0.081	-0.010	0.009	0.078	-0.001
Dollar volume	**Total**	129,314,442	200,594,040	169,581,962	62,578,907	102,898,702	85,626,677	233,441,454	368,801,854	309,270,232
	Small Trades	50,917,348	70,133,182	63,184,900	6,033,261	9,090,136	7,826,783	120,949,691	175,234,588	155,292,022
	Medium Trades	15,780,422	22,928,399	19,467,506	5,551,280	8,656,059	7,462,832	31,740,886	47,501,927	39,441,378
	Large Trades	62,616,671	107,532,459	86,929,556	50,994,367	85,152,508	70,337,063	80,750,877	146,065,339	114,536,832
Change in Proportion of Volume	**Small Trades**	0.041	-0.096	0.064	0.050	-0.101	0.065	0.028	-0.086	0.062
	Medium Trades	-0.001	-0.004	0.004	0.001	-0.011	0.009	-0.004	0.008	-0.004
	Large Trades	-0.041	0.099	-0.069	-0.052	0.111	-0.075	-0.024	0.079	-0.058
Change in Proportion of Imbalance	**Small Trades**	0.022	-0.044	0.031	0.024	-0.048	0.035	0.018	-0.037	0.024
	Medium Trades	-0.002	0.001	0.003	-0.003	0.002	0.002	-0.001	0.000	0.003
	Large Trades	-0.023	0.058	-0.038	-0.028	0.066	-0.043	-0.015	0.043	-0.030
Price Impact *t-stat*	**Small Trades**	0.1325	0.167	0.124	0.07	0.0934	0.073	0.175	0.273	0.185
		30.14	*39.32*	*43.96*	*21.12*	*29.07*	*32.24*	*28.76*	*41.53*	*47.94*
	Medium Trades	0.006	0.0114	0.0102	0.001	0.008	0.00243	0.009	0.013	0.016
		4.82	*5.33*	*8.34*	*0.41*	*2.89*	*1.17*	*3.42*	*4.15*	*8.65*
	Large Trades	0.002	0.013	0.005	-0.003	0.01	0.0011	0.011	0.013	0.009
		4.32	*6.12*	*7.31*	*-2.18*	*4.87*	*0.67*	*6.98*	*5.67*	*5.87*

Panel B – Decrease

		1993-2013			1993-2001			2002-2013		
		-1	**0**	**1**	**-1**	**0**	**1**	**-1**	**0**	**1**
Return		0.015	-0.048	0.005	0.020	-0.053	0.002	0.004	-0.036	0.013
Dollar volume	**Total**	113,831,764	35,269,466	47,664,477	69,612,253	21,417,758	30,344,181	208,998,373	66,865,588	87,597,995
	Small Trades	38,892,717	14,044,025	17,689,408	8,730,169	3,671,453	4,638,925	103,806,762	37,704,142	47,778,485
	Medium Trades	12,942,689	4,286,100	5,594,129	7,199,348	2,633,428	3,464,541	25,303,165	8,055,891	10,504,087
	Large Trades	61,996,359	16,939,341	24,380,940	53,682,735	15,112,877	22,240,716	79,888,446	21,105,556	29,315,422
Change in Proportion of Volume	**Small Trades**	-0.051	0.130	-0.063	-0.049	0.136	-0.066	-0.055	0.117	-0.055
	Medium Trades	-0.006	0.003	-0.001	-0.010	0.009	-0.004	0.002	-0.011	0.006
	Large Trades	0.058	-0.133	0.063	0.059	-0.145	0.070	0.054	-0.106	0.049
Change in Proportion of Imbalance	**Small Trades**	-0.029	0.086	-0.047	-0.021	0.087	-0.049	-0.044	0.083	-0.040
	Medium Trades	-0.002	-0.002	0.007	-0.003	0.000	0.006	-0.002	-0.007	0.009
	Large Trades	0.041	-0.077	0.045	0.042	-0.079	0.046	0.039	-0.071	0.040
Price Impact *t-stat*	**Small Trades**	0.152	0.01	0.155	0.08	0.066	0.091	0.202	0.133	0.232
		11.78	*31.43*	*35.33*	*16.98*	*27.53*	*29.86*	*19.07*	*28.67*	*33.81*
	Medium Trades	0.007	0.0063	0.01	0.007	-0.001	0.006	0.007	0.01	0.011
		1.69	*2.23*	*3.41*	*1.85*	*-0.52*	*2.09*	*1.45*	*4.24*	*3.87*
	Large Trades	0.01	0.003	0.001	0.011	-0.004	-0.003	0.01	0.006	0.011
		2.8	*0.87*	*1.25*	*2.87*	*-2.11*	*-1.43*	*2.32*	*2.86*	*4.18*

References

Barber, Brad, Terrance Odean, and Ning Zhu, 2009. Do retail trades move markets?, *Review of Financial Studies* 21, 151–186.

Barclay, Michael, and Jerold Warner, 1993. Stealth trading and volatility: Which trades move prices?, *Journal of Financial Economics* 34, 281–305.

Bessembinder, Hendrik, 2003. Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis* 38, 747-777.

Bessembinder, Hendrik, and Herbert Kaufman, 1997. A cross-exchange comparison of execution costs and information flow for NYSE-listed stocks, *Journal of Financial Economics* 46, 293-319.

Boehmer, Ekkehart, Gideon Saar, and Lei Yu, 2005. Lifting the veil: An analysis of pre-trade transparency at the NYSE, *Journal of Finance* 60, 783–815.

Chakrabarty, Bidisha, Bingguang Li, Vanthuan Nguyen, and Robert Van Ness, 2007. Trade classification algorithms for electronic communications network trades, *Journal of Banking and Finance* 31, 3806-3821.

Chakrabarty, Bidisha, Pamela Moulton, and Andriy Shkilko, 2012. Short sales, long sales, and the Lee-Ready classification algorithm revisited, *Journal of Financial Markets* 15, 467-491.

Chakravarty, Sugato, 2001. Stealth trading: Which traders' trades move stock prices?, *Journal of Financial Economics* 61, 289-307.

Chakravarty, Sugato, Venkatesh Panchapagesan, and Robert Wood, 2005. Did decimalization hurt institutional investors?, *Journal of Financial Markets* 8, 400–420

Chan, Kalok, Wai-Ming Fong, 2000. Trade size, order imbalance, and the volatility–volume relation, *Journal of Financial Economics* 57, 247–273.

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

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

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2011. Recent trends in trading activity and market quality, *Journal of Financial Economics* 101, 243-263.

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

Easley, David, Nicholas Kiefer, and Maureen O'Hara, 1997. The information content of the trading process, *Journal of Empirical Finance* 4, 159-186.

Easley, David and Mureen O’Hara, 1987. Price, Trade size and information in securities markets, *Journal of Financial Economics* 19, 69-90.

Ellis, Katrina, Roni Michaely, and Maureen O’Hara, 2000. The accuracy of trade classification rules: Evidence from NASDAQ, *Journal of Financial and Quantitative Analysis* 35, 529-551.

Fama, Eugene, and James MacBeth, 1973. Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.

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

Goldstein, Michael and Kenneth Kavajecz, 2000. Eighths, sixteenths, and market depth: Changes in tick size and liquidity provision on the NYSE, *Journal of Financial Economics* 56, 125–149.

Harford, Jarrad, and Aditya Kaul, 2005. Correlated order flow: Pervasiveness, sources and pricing effects, *Journal of Financial and Quantitative Analysis* 40, 29-55.

Hasbrouck, Joel, 1991. Measuring the information content of stock trades, *Journal of Finance* 46, 179-207.

Hendershott, Terrence, Charles Jones, and Albert Menkveld, 2011. Does algorithmic trading improve liquidity?, *Journal of Finance* 66, 1–33.

Hendershott, Terrence, and Pamela Moulton, 2011. Automation, speed, and stock market quality: The NYSE's Hybrid, *Journal of Financial Markets* 14, 568-604.

Heston, Steven, Robert Korajczyk, and Ronnie Sadka. 2010. Intraday patterns in the cross-section of stock returns, *Journal of Finance* 65, 1369-1407.

Holden, Craig, and Stacey Jacobsen, 2014. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *Journal of Finance* 69, 1747-1785.

Jain, Pankaj K, 2005. Financial market design and the equity premium: Electronic versus floor trading, *Journal of Finance* 60, 2955-2985.

Jones, Charles, Gautam Kaul, and Marc Lipson, 1994. Transactions, volume, and volatility, *Review of Financial Studies* 7, 631-651.

Kaul, Gautam, and M. Nimalendran, 1990. Price reversals: Bid-ask errors or market overreaction?, *Journal of Financial Economics 28*, 67-93.

Keim, Donald, and Ananth Madhavan, 1996. The upstairs market for large-block transactions: Analysis and measurement of price effects, *Review of Financial Studies* 9, 1-36.

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

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

O'Hara, Maureen, Chen Yao, and Mao Ye, 2014. What's not there: Odd-lots and market data, *Journal of Finance* 69, 2199-2236.

Petersen, Mitchell, 2009. Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

Roll, Richard, 1984. A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, 1127-1139.