Order characteristics and the sources of commonality in prices and liquidity

Shane A. Corwin

Department of Finance
Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556
scorwin@nd.edu

Marc L. Lipson

Darden Graduate School of Business Administration
University of Virginia
Charlottesville, VA 22906
mlipson@virginia.edu

forthcoming, *Journal of Financial Markets*

* We thank Bruce Lehmann (the editor) and an anonymous referee for their comments and suggestions. We also thank Rob Battalio, Jay Coughenour, Jeff Harris, Joel Hasbrouck, and seminar participants at the University of Delaware for helpful comments. Any remaining errors are the responsibility of the authors.

Order characteristics and the sources of commonality in prices and liquidity

Abstract

Using electronic order flow data for a sample of NYSE-listed stocks, we examine the relative importance of program traders, institutional traders, retail traders, and exchange members in driving commonality in order flow, returns, and liquidity. Using principal components analysis, we find that program trades and other institutional trades are the primary drivers of commonality in order flow and that these two order flow factors are significantly related to returns. Our results suggest that commonality is driven by the correlated trading decisions of professional traders, as executed through program trades, and not by correlated trading among retail traders.

JEL classification: G10; G11; G12; G20

Keywords: Commonality; Order flow; Liquidity; Program trading

Recently, a growing body of research has focused attention on common cross-security effects in order flow and liquidity and on the relation between these common factors and returns.[1] Despite this attention, little is known about the sources of these common effects or what drives the relation to returns. Prior studies suggest that commonality may reflect correlated trading decisions within specific groups of traders, such as retail investors or index-based traders. However, empirical evidence related to trader type is limited.[2]

Using a unique dataset consisting of all electronic order flow from November 1997 through February 1998 for a size-stratified sample of 100 NYSE-listed stocks, we address several unresolved issues related to the magnitude and sources of commonality. First, we test whether significant common components exist in the intraday order flows of program traders, non-program institutional traders, retail traders, and exchange members. Importantly, by using principal components analysis, we are able to extract and analyze common factors without assigning any particular role to the market portfolio. Second, we examine the relative importance of these trader groups for explaining order flow commonality and the relation between order flow commonality and the common components in returns and liquidity. While some evidence exists with respect to retail traders and mutual funds, the relative strength of commonality for these groups has not been addressed and no direct evidence exists with respect to commonality in program or member trading. Unlike prior studies, we are able to directly identify trader type and are not limited to proxies based on trade size, firm size, or S&P index inclusion. Third, we

---

[1] See, for example, Chordia, Roll, and Subrahmanyam (2000), Hasbrouck and Seppi (2001), and Korajczyk and Sadka (2008).

[2] Behavioral studies suggest that common factors may be driven by style-based trading [e.g., Barberis and Shleifer (2003)] or habitat investing [e.g., Lee, Shleifer, and Thaler (1991)]. Empirical studies tend to support these explanations using data for particular subsets of securities such as closed-end funds and securities added to the S&P 500, and for particular subsets of traders such as retail traders and mutual funds. This evidence is discussed in more detail in Section 1.

provide evidence on the importance of firm size and other firm characteristics in driving commonality in returns, order flow, and liquidity.

We find significant common factors in numerous measures of order flow based on both submissions and executions of market and limit orders. For example, the first principal component explains approximately 8% of the variation in signed market plus limit order submissions. More important, when we decompose order flow by trader type, we find that commonality is strongest for program trades. In particular, we find that program trading accounts for as much as 86% of the common factor in order flow, while other institutional trading accounts for the remaining 14%. Though weak common factors appear to exist in both retail and member order flow, neither of these appears to significantly affect aggregate order flow commonality.

We find that order flow common factors are related to commonality in returns and that this relation is driven primarily by program traders. Across all stocks and intraday periods, the common factor in returns explains approximately 8.4% of return variation. Of this common component, roughly 51% can be attributed to commonality in program trading and an additional 8% can be attributed to commonality in other institutional order flow. After controlling for these components, retail trading and member trading have little incremental impact on returns. We also find that the importance of order flow common factors differs across firm size categories. After controlling for own order flow, both commonality in program order flow and commonality in other institutional order flow help explain returns on large firms. In contrast, small firm returns are driven primarily by own order flow and are only marginally related to order flow common factors. Our findings are consistent with the predominant role of program traders and institutional investors in the trading of large stocks. The results also suggest that commonality related to retail investor sentiment has little incremental explanatory power for returns in our sample.

While principal components analysis assigns no particular economic meaning to the common factors, the cross-section of factor loadings provides some insight into their economic interpretation. For both returns and order flow, our results suggest that the first principal component is related to market-wide variation whereas the second principal component reflects differences between large and small firms. We note that the second common factor is analogous to the SMB factor analyzed in three-factor asset pricing models and is consistent with style-based trading explanations of commonality. This finding also suggests that commonality studies based solely on a weighted-average market factor or solely on large firms may fail to capture important systematic components in order flow and liquidity.

Consistent with prior research, we find that common factors explain only a small fraction of the variation in quoted bid-ask spreads. For example, the first principal component explains only 2%-3% of that variation. Commonality is stronger for liquidity measures derived from limit order book depth. For example, the first principal component explains 7%-10% of the variation in limit book depth. Common liquidity factors are also stronger for small firms than for large firms. These results suggest that common factors in liquidity may be more important than suggested by analyses that focus only on quoted spreads or only on large stocks. After controlling for security-specific characteristics, we find that common order flow components tend to be negatively related to individual security spreads, regardless of trader type. These results suggest that common components in order flow may reflect liquidity trading rather than informed trading. We also find that individual security spreads are significantly related to commonality in limit versus market order use, as shown in Domowitz, Hansch, and Wang (2005).

Our analysis is closely related to Hasbrouck and Seppi (2001) and Korajczyk and Sadka

(2008). Hasbrouck and Seppi use principal components analysis to examine commonality in intraday trades, returns, and liquidity for the Dow 30 stocks. They identify a significant common factor in signed trades and find that this factor explains approximately two-thirds of the common factor in returns. Like Chordia, Roll, and Subrahmanyam (2000), they also find evidence of commonality in liquidity, but conclude that the economic significance of these common factors is small. Korajczyk and Sadka use asymptotic principal components to analyze systematic liquidity across eight alternative monthly liquidity measures. They find that across-measure systematic liquidity is a priced factor, while measure-specific (within-measure) systematic liquidity adds no incremental pricing information. Using intraday order flow data for a broad sample of stocks, we provide new evidence on the sources of this documented commonality. Following the principal components analysis of Hasbrouck and Seppi, we find that multiple common factors exist in both returns and order flow. In addition, our results show that these common factors are driven primarily by program trading and that other classes of traders play only a small role. Given the growing importance of program trading, these results suggest that commonality in order flow and the effects of this commonality on returns are likely to increase over time.

Our findings are also related to two recent papers analyzing the role of institutional investors in explaining commonality. Using data at 15-minute intervals, Harford and Kaul (2005) find that commonality in signed trades and returns is stronger for S&P stocks than non-S&P stocks, is stronger in 1996 than 1986, and is stronger at the end of the trading day. They conclude that common effects are driven primarily by indexing and are not pervasive. Like Hasbrouck and Seppi (2001), Harford and Kaul find only weak evidence of commonality in trading costs. Kamara, Lou, and Sadka (2008) use a daily liquidity measure based on Amihud (2002) to analyze changes in the cross-sectional variation in liquidity commonality from 1963 through

2005. They find that the sensitivity of individual stock liquidity to market-wide liquidity has increased over time for large stocks and decreased over time for small stocks. They provide evidence that this divergence is related to patterns in institutional ownership and tie these effects to similar patterns in return commonality. We contribute to these studies along several dimensions. First, as noted earlier, our trader type identification allows us to directly analyze the strength of trader-specific common factors and the relative importance of these factors in explaining variation in returns and liquidity. We identify strong commonality in order flow from program traders, but only weak commonality in order flow from non-program institutional traders and retail traders. These findings suggest that Harford and Kaul's results related to S&P inclusion are driven primarily by program traders. The results may also provide at least a partial explanation for the time-series patterns analyzed in Kamara, Lou, and Sadka (2008). Second, our order flow data allow us to analyze commonality in both order submissions and executions, and to analyze commonality in liquidity measures based on limit book depth beyond the posted bid-ask quotes. Our findings suggest that commonality is evident in both submissions and executions, and is stronger for limit-book liquidity measures than for quoted spreads. Finally, we provide evidence consistent with an additional size-related common factor in order flow and returns.

The remainder of the paper is organized as follows. In Section 1, we provide a review of prior literature related to the potential determinants of commonality. Section 2 describes the sample and the characteristics of the order flow data. Section 3 describes the principal components and canonical correlation methodology. In Section 4, we examine the importance of common factors in returns and order flows. We also assess the relative importance of program traders, institutional traders, retail traders, and exchange members in explaining common factors in order flow and returns. In Section 5, we test for common factors in liquidity and analyze the effects on liquidity of order flow from various types of traders. Section 6 concludes.

1. The determinants of commonality

Commonality in returns can be driven by many factors. The traditional view is that return commonality is driven by comovements in underlying fundamentals. However, empirical evidence suggests that the observed levels of return commonality exceed that which can be explained by fundamentals. For example, Fama and French (1995) identify common factors in the returns of both small stocks and value stocks and find that these factors cannot be fully explained by commonality in the underlying earnings. Froot and Dabora (1999) study the "twin" stocks of Royal Dutch, traded primarily in the United States, and Shell, traded primarily in the United Kingdom. Although these stocks represent claims on the same underlying cash flows, Froot and Dabora show that Royal Dutch stock is more highly correlated with U.S. market movements and Shell stock is more highly correlated with U.K. market movements.

Several alternative explanations for return commonality have been proposed in the literature. These explanations, as discussed in Barberis, Shleifer, and Wurgler (2005), generally rely on some type of market frictions and can be classified into three groups. First, commonality could be driven by style investing. Barberis and Shleifer (2003) develop a model of asset prices where some investors categorize assets into different styles, moving funds between style categories based on relative performance. This type of style investing leads to common factors in the returns (and order flow) of securities within the same style and reduces the correlations between stocks in different styles. Barberis and Shleifer suggest that style investing provides a reasonable explanation for many empirical findings related to commonality, such as commonality in the returns of value stocks.

A second explanation is based on habitat investing, as described in Lee, Shleifer, and

Thaler (1991). They argue that some securities may be held by only particular subsets of investors. For example, small stocks and closed-end funds tend to be held primarily by individual investors. As the sentiment or risk preferences of these investors change, their resulting trades may lead to common factors in the returns (and order flow) of the securities they hold. Consistent with this theory, Lee, Shleifer, and Thaler find evidence of commonality in the prices of closed-end funds and show that these funds co-move with small-cap stocks even when the funds hold primarily large-cap stocks. Bodurtha, Kim, and Lee (1995) provide related evidence for closed-end country funds. They find that the prices on these funds tend to co-move with the U.S. market and interpret this as evidence of a U.S.-specific risk or sentiment factor.

A third potential explanation for commonality is related to differences in the speed of information diffusion. Here, the prices of different securities are expected to react to information at different rates. For example, some securities may react to market or industry information immediately, while others react with a one hour or one day lag. This leads to commonality in the returns of securities with similar rates of information diffusion. Barberis, Shleifer, and Wurgler (2005) provide evidence related to this and other potential explanations of commonality using data on S&P index changes. Their analysis builds on that of Vijh (1994), who shows that betas relative to a value-weighted market portfolio increase for securities added to the S&P 500 index. Barberis, Shleifer, and Wurgler extend this analysis to show that betas relative to the S&P portfolio increase, while betas relative to non-S&P stocks decrease following S&P additions. The opposite results are found following S&P deletions. These results provide support for friction- or sentiment-based explanations of commonality but are not consistent with fundamentals-based explanations. When they decompose these effects, Barberis, Shleifer, and Wurgler find that at least a portion of the effect is driven by differences in information diffusion across stocks.

Notably, style and habitat-based explanations suggest that common factors reflect correlated trading decisions within specific groups of traders. For example, sentiment effects may be more evident among retail traders, while style effects are likely to be important in the trading activity of institutions such as mutual funds or program traders. Existing evidence related to common factors in retail trading is somewhat mixed. Using monthly trade data from individual investor retail brokerage accounts during the period from 1991 to 1996, Kumar and Lee (2006) find a strong positive correlation in signed trading activity across different portfolios of stocks and different groups of investors. They also show that portfolio-level measures of signed retail trading have incremental explanatory power for monthly returns. In contrast, Kaniel, Saar, and Titman (2008) study retail trading using daily audit trail data for NYSE-listed securities from 2000 through 2002 and find only weak evidence of commonality in retail trades. They report that the first principal component in signed retail trades accounts for only 1.33% of the total variation in signed trading. The effects of commonality in institutional trading are analyzed by Harford and Kaul (2005), who compare return and order flow commonality in S&P and non-S&P stocks.[3] They find that commonality is most pronounced for index stocks and conclude that indexing is the primary source of commonality.[4] We contribute to the prior literature by providing direct evidence on the relation between commonality and correlated

[3] While only limited evidence exists with respect to cross-security effects, the effects of institutional trades on individual security prices are well documented. See, for example, Chan and Lakonishok (1993, 1995, 1997), Keim and Madhavan (1997), and Jones and Lipson (1999). Warther (1995) and Edelen and Warner (2001) provide additional evidence at the aggregate level. Warther finds that monthly market returns are highly correlated with unexpected aggregate cash flows into mutual funds. Edelen and Warner also report a positive correlation at the daily level, but note that fund flows explain only about 3% of daily variation in market returns. Harris, Sofianos, and Shapiro (1994) analyze price effects related to program trades.

[4] Harford and Kaul (2005) are not able to directly identify indexing trades. However, they hypothesize that indexing should be strongest at the end of the trading day and should be stronger in 1996 than 1986. They also compare commonality effects before and after S&P 500 index additions. In related work, Hughen and McDonald (2006) analyze comovements within different security classes and different trade size categories. They report evidence of common factors for both small- and medium-sized trades, but find that only comovements in medium-sized trades affect returns.

trading decisions within specific trader groups. We also analyze the relative importance of various types of traders in driving commonality in order flow, returns, and liquidity.

In the context of liquidity, commonality may be driven by factors that affect either the supply of or demand for liquidity within groups of stocks.[5] For example, Coughenour and Saad (2004) identify a specialist firm factor in liquidity for NYSE-listed stocks, suggesting that comovements may be driven by common factors in the costs of providing liquidity. However, liquidity provision is not limited to market-makers, as investors on both the buy and sell sides can choose to either provide liquidity in the form of limit orders or take liquidity in the form of market orders. As a result, commonality in liquidity may be driven by both the direction of trade and the market versus limit order decision. This idea is formalized by Domowitz, Hansch, and Wang (2005), who show that return commonality is driven by order flow while liquidity commonality is driven by order type.[6] Liquidity supply and demand may also be affected by habitat and style-based trading. For example, inclusion in the S&P 500 index could increase both the demand for and supply of liquidity from institutional investors. Similarly, a shift in individual investor sentiment could lead to changes in liquidity supply or demand for the subset of stocks traded by these investors. Our analysis using order flow data allows us to examine the importance of supply and demand effects within specific subsets of traders.

[5] Chordia, Roll, and Subrahmanyam (2000) also suggest that commonality in liquidity may reflect industry-specific liquidity and information shocks.

[6] Bloomfield, O'Hara, and Saar (2005) examine market versus limit order choice in an experimental setting. They find that informed traders use more limit orders than liquidity traders and liquidity provision changes as prices adjust to information. Specifically, informed traders tend to take liquidity when the value of their information is high and shift to providing liquidity as prices adjust. In this context, commonality in relative market and limit order use could be driven by informed traders, who possess information relevant to an entire industry or sector, or by common time variation in the rate of price adjustment to new information.

2. Sample and order flow characteristics

The primary data source for our analysis is NYSE system order data for the period from November 1997 through February 1998. The sample period is determined primarily by data limitations. However, this time period also has a significant advantage over more recent periods. During this time period, the NYSE not only captured the vast majority of order flow in NYSE-listed stocks, but also required order flow to be categorized by trader type. This allows us to carefully identify the common factors in various order flow categories. The sample includes 100 NYSE-listed common stocks selected as follows. First, all domestic NYSE common stocks were ranked according to market capitalization as October 31, 1997.[7] Securities were then sorted into deciles based on market capitalization and ten securities were selected randomly from each decile. For each security, we supplement the NYSE system order data with trade and quote data from TAQ, price and market capitalization data from CRSP, book values from Compustat, and institutional holdings from Thompson Financial.

Because trading activity and liquidity vary significantly by firm size, we provide results for the full sample and for subsamples based on market capitalization. We define small stocks as those in market capitalization deciles 1-3, midcap stocks as those in deciles 4-7, and large stocks as those in deciles 8-10. This breakdown also allows for comparability to prior research.

Summary statistics for the sample are provided in Panel A of Table 1. Market capitalization and closing price are measured at the start of the sample period. Institutional holdings and book value are measured at the end of 1997. All other variables are averaged across all trading days in the sample period. Market capitalization averages \$5.6 billion in the full sample and ranges from \$304 million for small stocks to \$16.5 billion for large stocks. As

[7] Stocks are excluded if they cannot be identified in CRSP or if they change listing venue or delist during the sample period.

expected, large stocks have higher prices and more trading activity than small stocks. For example, average daily share volume ranges from 61,000 for small firms to 814,000 for large firms. The standard deviation of daily returns also tends to be lower for large firms (1.9%) than small firms (2.2%), but equality of means across size groups is not rejected for this variable.[8] The book-to-market ratio, as defined in Fama and French (1993), averages 0.49 for the full sample and ranges from 0.35 for small firms to 0.62 for large firms. The proportion of shares held by institutions averages 55.6% and ranges from 44.3% for large firms to 64.4% for small firms.

The system order data used in our analysis include all market and limit order submissions, executions, and cancellations through the NYSE's SuperDOT system during the sample period. These data are similar to the order flow data in the NYSE's TORQ database, as described in Hasbrouck (1992). The data contain detailed information on all SuperDOT orders, including submission time, number of shares submitted, order side (buy, sell, short sell, etc.), order type (market or limit), order conditions (day order, good-till-cancelled, etc.), limit price (if applicable), and account type. If any portion of the order is executed or cancelled, the data also include the time of execution or cancellation, the number of shares involved, the number of shares remaining in the original order, and the execution price (if applicable). If an order results in multiple executions or cancellations, data are provided for each separate event.

These detailed order data have several advantages over execution data such as that provided in TAQ. First, we can observe actual order flow in addition to executions. This allows

[8] For comparison, we note that CRSP identifies 1,660 NYSE-listed domestic common stocks with stock prices between $1 and $500 as of December 31, 1997. These firms have an average market capitalization of $4.8 billion, an average price of $33.04, and an average daily volume of 299,000 shares. Hasbrouck and Seppi (2001) analyze the 30 stocks of the Dow Jones Industrial average across all trading days in 1994. The average firm in their sample trades 608 times per day, with an average price of $51, and a daily return standard deviation of 1.5%. These values are similar to those reported in Table 1 for the large stock subsample.

us to test whether order flow itself is informative above and beyond executions. Second, orders are signed, eliminating the need to use quote-based algorithms to sign trades. For example, unlike quote-based algorithms, we can observe the direction of all orders including those executed at the open and those executed at the quote midpoint.[9] Third, orders are identified by account type, allowing us to separate order flow from program traders, non-program institutional traders (hereafter “institutional traders”), retail traders, and exchange members. Finally, the data allow us to reconstruct the limit order book, making it possible to address questions of liquidity using limit book information in addition to the quoted bid-ask spread.

While these data have several advantages, they also have two important limitations. First, the data include only orders entered through the SuperDOT system. Orders handled manually by floor brokers and specialists are not included. The 1998 NYSE Fact Book states that “the vast majority of orders, representing about 44% of the volume” are routed directory to the specialist using the SuperDOT system. Consistent with this, across all securities and trading days in our sample, the ratio of shares executed through SuperDOT to twice NYSE share volume is 49.5%.[10] Notably, this data limitation is unlikely to affect our principal results related to program and retail orders, as these orders are not likely to be handled by floor brokers. Nevertheless, we include signed TAQ trades to account for any incremental explanatory power due to orders handled on the trading floor.

The second limitation of the data is the time period covered. On the one hand, the vast majority of order flow for NYSE-listed stocks went through the NYSE trading floor during this time period, allowing us to better isolate the effects of specific types of traders. On the other

[9] While many prior studies refer to order flow and order imbalances, these studies generally use signed TAQ trades as a proxy for order flow.

[10] In comparison, Sofianos and Werner (2000) find that SuperDOT accounted for an average of 45% of total buy plus sell volume during the first two months of 1997, while specialists accounted for 11% and floor brokers 44%.

hand, we are unable to study the effects on commonality of several important market structure changes, including the implementation of Reg-NMS and the rapid growth in program and algorithmic trading during more recent periods.[11] We leave a more detailed analysis of these effects to future research.

The characteristics of the order flow data are illustrated in Figure 1. Panels A and B plot the average volume of limit and market order submissions per 15-minute period by trader type. Results are provided for the full sample and for the subsamples of small, midcap, and large firms. On average, institutional traders account for the majority of submissions, for both limit and market orders. Not surprisingly, program traders account for a larger fraction of order flow in large stocks than in small stocks. In addition, program traders account for a larger fraction of limit orders than market orders. In contrast, retail traders play a more active role in small stocks than large stocks and represent a larger fraction of market orders than limit orders. Member trading is most evident in the largest stocks, especially for limit orders. Although not shown, results based on executions are similar.

Panel C of Figure 1 plots the proportion of shares submitted as limit orders. Again, the results are broken down by trader type and firm size. On average, limit orders account for approximately 60%-70% of order submission volume. However, this pattern differs markedly by trader type and firm size. For submissions by program traders, institutions, and members, the limit order ratio ranges from approximately 60% for small firms to over 80% for large firms. In contrast, the limit order ratio for retail traders ranges from only 30% for small firms to over 50% for large firms. Again, results for executions follow the same general pattern and are not shown.

[11] According to the NYSE's Market Data, the NYSE captured 83% of consolidated volume in NYSE-listed securities during 1998. In comparison, the NYSE Group's share of consolidated volume in 2009 was less than 40%. Program trading as a fraction of total NYSE volume increased from 18% in 1998 to approximately 30% in 2009.

Overall, the results in Figure 1 suggest that order flow characteristics differ significantly both across firms and by trader type.

To further assess the state of liquidity, we reconstruct the limit order book at 15-minute intervals following the methodology of Kavajecz (1999).[12] To illustrate the characteristics of the limit book, Figure 2 plots cumulative limit book depth at various increments from the quote midpoint. Results are provided for the full sample and the subsamples of large, midcap, and small firms. The limit order book tends to exhibit a substantial sell imbalance, with total bid depth averaging approximately 25,000 shares and total ask depth averaging over 40,000 shares. This depth imbalance is evident both near the quotes and away from the quotes and for both small and large firms. Comparing depth across firm size categories, we see that total limit book depth is quite similar for small and midcap stocks. However, these groups differ markedly from large firms. Total bid depth averages 17,000 to 18,000 shares for small and midcap firms, compared to nearly 60,000 shares for large firms. Total ask depth averages approximately 20,000 shares for small and midcap firms, compared to over 90,000 shares for large firms.

To define liquidity, we estimate quoted bid-ask spreads and two measures of execution costs derived from information in the limit order book. The limit spread is defined as the difference between the best ask and bid prices in the limit order book. Like the quoted spread, this measure reflects potential execution costs for small orders. To reflect potential execution costs for larger orders, we incorporate additional information from the limit order book. The

[12] This methodology involves three steps. First, we construct the initial limit book using all orders executed or cancelled during the sample period with submission dates prior to the sample period. Second, at each 15-minute interval, we incorporate the set of limit orders submitted prior to the date and time of interest. Third, we use execution and cancellation records to remove any orders that were executed or canceled and update any orders that were partially executed or canceled prior to the date and time of interest. "Day" orders are dropped at the end of each day. Execution and cancellation records are matched to original submissions using submission time and date, along with all available order identifiers and characteristics. To minimize errors, we also delete limit buy (sell) orders at prices higher (lower) that the posted ask (bid) quote.

1,000-share spread is defined as the difference between the 1,000-share ask and the 1,000-share bid, where the 1,000-share ask (bid) is the price one would have to go up (down) to in order to reach 1,000-shares of cumulative depth in the limit order book.

Summary statistics for the three liquidity measures are provided in Panel B of Table 1. Consistent with the higher trading activity and lower volatility in the large firm subsample, these firms exhibit significantly lower bid-ask spreads. Quoted spreads average 0.63% in the full sample and range from an average of 0.30% for large firms to 1.01% for small firms. The limit order book spread averages approximately three times the quoted spread and the 1,000-share spread averages three to five times the quoted spread, with the difference being largest for small stocks.[13]

3. Methodology for measuring commonality

Following Hasbrouck and Seppi (2001), we use principal components analysis to extract and analyze common factors in returns, order flow, and liquidity. This statistical method for identifying common factors relies solely on the variance-covariance matrix and, unlike the "market model" framework of Chordia, Roll, and Subrahmanyam (2000), assigns no special role to the market portfolio. This allows us to analyze both the relative importance of common factors in various order flow categories and the relations among common factors in order flow, returns, and liquidity. It also allows us to test for the presence of multiple common factors. In the discussion to follow, we provide some economic interpretation of these extracted common factors in relation to market-wide effects and other economic factors.

[13] We note that these limit book spread measures reflect only displayed liquidity. Actual execution costs for large orders are likely to differ substantially from these values when floor traders and specialists actively participate in trading. To limit outliers, we set the maximum 1,000-share bid (ask) to the quoted bid minus $1 (quoted ask plus $1). Although not shown, conclusions based on a 5,000-share spread measure are similar. The 5,000-share spread averages five-to-eight times the quoted spread.

Principal components analysis explains the variance-covariance structure of the underlying data using linear combinations of the original variables. These linear combinations are chosen in order to maximize explanatory power.[14] Let x_{it} be a variable observed for security *i* at time *t*. For example, in our analysis, x_{it} could represent measures of return, order flow, or liquidity. Given *N* securities and *T* periods, let $\boldsymbol{X_1}$, $\boldsymbol{X_2}$, …, $\boldsymbol{X_N}$ be vectors of length *T* such that $\boldsymbol{X_i}'=[x_{i1}, x_{i2}, \ldots, x_{iT}]$, let $\boldsymbol{X}$ be the matrix $[\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_N}]$, and let $\sum$ be the covariance matrix of $\boldsymbol{X}$. Consider a linear combination of these variables defined as:

$$\boldsymbol{Y_j} = \boldsymbol{\gamma_j}'\mathbf{X} = \boldsymbol{\gamma_{1j}}\mathbf{X_1} + \boldsymbol{\gamma_{2j}}\mathbf{X_2} + \ldots + \boldsymbol{\gamma_{Nj}}\mathbf{X_N}. \tag{1}$$

The variance of $\boldsymbol{Y_j}$ is then defined as $\boldsymbol{\gamma_j}'\sum\boldsymbol{\gamma_j}$.

If $\boldsymbol{\gamma_j}$ is chosen to maximize the variance of $\boldsymbol{Y_j}$, then $\boldsymbol{Y_j}$ is referred to as the first principal component. The second principal component is the linear combination that maximizes the variance of $\boldsymbol{Y_j}$ subject to the condition that it be uncorrelated with the first principal component. As many as *N* principal components can be extracted in this manner conditional on each being uncorrelated with all previously extracted principal components. Together, the *N* principal components provide the same information as the original *N* variables. In other words, $\sum_{i=1}^{N}\sigma_i^2 = \sum_{j=1}^{N}Var(Y_j)$. We also note that the variance of $\boldsymbol{Y_j}$ can be arbitrarily increased by multiplying $\boldsymbol{\gamma_j}$ by a constant. To account for this, the $\boldsymbol{\gamma_j}$ are chosen to be of unit length such that $\boldsymbol{\gamma_j}'\boldsymbol{\gamma_j}=1$.

A useful feature of principal components analysis is that the principal components can be interpreted as functions of the eigenvalues and eigenvectors of $\sum$. In particular, the variance of the first principal component, $\lambda_1=\mathrm{Var}(\boldsymbol{Y_1})$, equals the first eigenvalue of $\sum$ and the coefficient vector, $\boldsymbol{\gamma_1}$,equals the first eigenvector. Similar notation can be used for each of the *N* principal

[14] For a more complete discussion of principal components analysis, see Johnson and Wichern (2007).

components. Throughout the paper, we report eigenvalues as a measure of the strength of commonality. We also examine the individual security loadings of the eigenvectors as a means of interpreting the common components.

Following Hasbrouck and Seppi (2001), we remove deterministic time-of-day effects by standardizing each of the variables based on security and time-of-day specific means and standard deviations.[15] Let x_{ipt} denote the observation of variable x for security i during intraday period p of day t. Then the standardized variable z is defined as:

$$z_{ipt} = \frac{x_{ipt} - \mu_{ip}}{\sigma_{ip}}, \tag{2}$$

where μ_{ip} and σ_{ip} are the mean and standard deviation across all days for security i during intraday period p. Standardizing the variables in this manner also aids in interpretation. Because the standardized variables have unit variance, we get:

$$\sum_{i=1}^{N} \sigma_i^2 = \sum_{j=1}^{N} Var(Y_j) = \sum_{j=1}^{N} \lambda_j = N. \tag{3}$$

In other words, the total variation equals the number of original variables (the number of securities in our context). In addition, the proportion of total variation explained by the j^{th} principal component equals:

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + \ldots + \lambda_N} = \frac{\lambda_j}{N}. \tag{4}$$

4. Commonality in returns and order flow

4.1. Commonality in returns and trading activity

We begin by analyzing the importance of common factors in returns and in trading measures from TAQ. This analysis is similar to that in Hasbrouck and Seppi (2001) and allows a

[15] Time of day effects are analyzed in McInish, Wood, and Ord (1985), McInish and Wood (1992), and Foster and Viswanathan (1993).

comparison with their results. A summary of the principal components analysis for these variables is provided in Table 2. Results for the full sample are provided in Panel A, with results for large and small securities in Panels B and C, respectively.[16] For informative purposes, the first two columns of the table show the mean and standard deviation of the original variable across all securities and all time periods. However, all principal components analyses are performed on the standardized variables. The remaining columns of the table report the first three eigenvalues (as a measure of the strength of commonality) and the cumulative proportion of total variation explained by the first three principal components. If there were no common components in the original variables, each eigenvalue would equal one and the first three principal components would explain $3/N$ of the total variation.

Strong evidence of common components is evident for all variables. For returns, the first eigenvalue is 8.43, suggesting that the first common factor explains 8.43% of total variation in the return variables. The second and third principal components are much smaller, but are still larger than one. This contrasts with the evidence in Hasbrouck and Seppi (2001) for the Dow 30 stocks and suggests that more than one common factor may be evident in returns. In total, the first three principal components explain approximately 11.7% of total variation in returns. We discuss the interpretation of these components in Section 4.4.

The results for trading activity also provide strong evidence of commonality. While evident for all measures of trading activity, commonality is strongest for total (unsigned) trades. Here the first eigenvalue is 8.14 and the first three principal components explain 13.4% of total variation. For trade imbalance, the first eigenvalue is 4.9 and the first three principal components explain 7.9% of total variation. Commonality also appears to be stronger for small trades (less

[16] To conserve space throughout the paper, we do not report separate results for the subsample of mid-cap securities (size deciles 4 through 7). The results for these securities support the conclusions based on the large and small firm subsamples.

than 500 shares) than for large trades (at least 10,000 shares). As noted by Hasbrouck and Seppi (2001), the weak commonality for block trades may reflect the mechanics of block trading where orders can be shopped in the upstairs market or worked on the trading floor. Though not shown, results based on share volume are similar.

Although the ability of principal components to identify common factors does not require any particular distributional assumptions, the significance of our eigenvalue estimates can be explored by considering the sampling distribution of the eigenvalues, λ_i, under multivariate normality.[17] If the original variables are drawn from a normal population, then the standard error for the eigenvalue is equal to $\lambda\sqrt{2/T}$. Given 80 trading days and 26 intraday periods, T=2,080 and the estimated eigenvalues will be approximately 32 times the magnitude of the standard errors under normality. For example, given that the first eigenvalue for returns is 8.43, the standard error under Normality would be approximately $\sqrt{2(8.43^2)/2{,}080} = 0.261.$ While these standard errors may be understated due to a violation of the normality assumption, they assist with the interpretation of the statistical significance of the eigenvalues.

While the results in Table 2 are generally consistent with the evidence in Hasbrouck and Seppi (2001), they differ in two important respects. First, the strength of the first common factor tends to be lower in our sample than in Hasbrouck and Seppi. For example, they report that the first principal components for returns and signed trades explain 21.1% and 11.2% of related variation, respectively. In comparison, these principal components explain 8.4% and 4.9% of related variation in our sample. Second, we find evidence of more than one common factor in both returns and trading activity. The second eigenvalue for returns is 1.82, compared to 1.04 in Hasbrouck and Seppi. For total trades, where the effect is even stronger, the second eigenvalue in

[17] See Hasbrouck and Seppi (2001) and Johnson and Wichern (2007) for more detailed discussions.

our sample is 3.08. One potential explanation for this difference in results is the role of midcap and small stocks in our sample. Hasbrouck and Seppi study only the 30 Dow stocks, while our sample includes 100 NYSE stocks across all size deciles. This allows us to capture potential commonalities across size groups.

Consistent with the discussion above, Panels B and C in Table 2 provide little evidence of more than one common factor when subsamples of large or small stocks are analyzed in isolation. For large stocks (Panel B), the results are very similar to the findings of Hasbrouck and Seppi (2001). The first eigenvalues for returns and signed trades are 5.96 and 3.99, respectively. This suggests that the first principal component for returns explains approximately 19.9% (5.96/30) of total return variation and the first three principal components explain approximately 27.1% ((5.96+1.11+1.05)/30).[18] Comparing Panels B and C, we see that commonality is strongest for large stocks. For small stocks, the first eigenvalue of returns is 2.01 and the first eigenvalue of signed trades is only 1.21. The weak commonality for small stocks may suggest that commonality is driven primarily by program traders or institutional traders, who tend to be more active in large securities. We address this issue in more detail below.

4.2. Commonality in order flow

We now turn our attention to commonality in order flow. Panel A of Table 3 summarizes the principal components analysis for order flow variables using the full sample of 100 securities. As in Table 2, we report means and standard deviations for all original variables and

[18] In the absence of any common components, each eigenvalue should equal 1 and account for a proportion 1/*N* of total variation. For the large and small stock subsamples, *N*=30, suggesting that the first three principal components will account for 3/30=10% of total variation in the absence of common components. In the full sample, *N*=100, suggesting that the first three principal components will account for 3/100=3.0% of total variation in the absence of common components.

principal components analysis is estimated using the standardized variables. Results are provided for both order submissions and order submission imbalances, using total orders, limit orders only, and market orders only. Although not reported, results based on order executions are similar to those reported for submissions.

Significant commonalities are evident for all order flow variables. For example, the first eigenvalue for total market plus limit order submission imbalance is 7.61 and the first three principal components for this variable explain approximately 12% of total variation. Again, based on the full sample, there appear to be at least two significant common components in the order flow measures. For both total submissions and submission imbalance, commonality appears stronger for limit orders than for market orders. There is also significant commonality in the relative use of limit and market orders. The first eigenvalue for the limit order ratio equals 5.81 and the first three principal components explain roughly 10.1% of total variation.

Results for large and small securities are provided in Panels B and C of Table 3, respectively. As in Table 2, common factors are strongest for large stocks and there is only weak evidence of multiple common factors. For large securities, the first eigenvalue for total submissions is 4.11 and the first eigenvalue for total submission imbalance is 3.92. While market order submission imbalance exhibits substantial commonality only for large securities (λ_1=4.29), limit order submission imbalance and relative use of limit orders exhibit strong commonality for all size categories. These findings are in stark contrast to the weak commonality evidence for small stocks based on TAQ trade variables and suggest that order flow data may provide valuable information with respect to commonality.

4.3. Determinants of commonality in order flow

Table 3 provides strong evidence of commonality in order flow for NYSE stocks. However, these results provide little guidance as to the sources of commonality. Prior research suggests that commonality may be driven by program trading or by institutional traders such as mutual funds. Commonality could also be driven by investor sentiment among retail traders. To test these possibilities, we estimate principal components for all order flow variables defined separately for program traders, institutional traders, retail traders, and exchange members. For each trader-specific order flow variable, Table 4 reports the total variation explained by the first principal component. Results based on aggregate order flow are reported for comparison.

In the full sample (Panel A in Table 4), it is clear that commonality is strongest for order flow from program traders. The first eigenvalue for total submissions is 11.01 and the first eigenvalue for total submission imbalance is 17.66. For comparison, the first eigenvalues for combined order flow imbalance and TAQ trade imbalance are 7.61 and 4.90, respectively. A single common component appears to explain 10%-20% of the variation in program trading order flow. Common components are also evident in order flow from institutional traders, retail traders, and exchange members. However, the results are much weaker than for program traders. For institutional traders, the first eigenvalues for total submissions and total submission imbalance are 2.96 and 3.05, respectively, suggesting that a single common factor explains approximately 3% of the variation in institutional order flow. For both retail and member order flow, the first common factor explains roughly 2% of total variation.

Results for the subsamples of large and small stocks are shown in Panels B and C, respectively, in Table 4. Commonality in program trading is evident for all categories, but again is strongest for large firms. For large firms, the first principal component explains approximately 27.9% of the variation in total program submissions and 26.2% of the variation in total program submission imbalance. The strongest common factor for large firms is related to market order submissions by program traders, while common factors for small firms are stronger for limit order submissions by program traders. For all size categories, a strong common factor is present in the relative use of market and limit orders by program traders. Among the remaining three trader types, commonality is strongest for institutional traders.

The results in Table 4 provide evidence of strong common factors in order flow from program traders and, to a lesser extent, in order flow from institutional traders. However, these common factors may be highly correlated. To examine the relative importance of the various trader types in explaining order flow commonality, we regress the first principal component from combined order flow on the trader-specific principal components. The regression takes the following form:

$$P_t^{CombinedOrderFlow} = \alpha + \beta_1 \cdot P_t^{Program} + \beta_2 \cdot P_t^{Institutional} + \beta_3 \cdot P_t^{Retail} + \beta_4 \cdot P_t^{Member} + \upsilon_t \,, \qquad (5)$$

where $P_t^{CombinedOrderFlow}$ is the first principal component from combined order flow and the remaining P_t are the first principal components from trader-specific order flow. We report results for two order flow variables: total market plus limit order submissions and total submission imbalance. Table 5 summarizes the R^2s from these regressions. Because commonality appears strongest for program trading and institutional trading, we report incremental R^2s from regressions where the explanatory variables are added sequentially starting with the program trading principal component, then the institutional trading component, the retail component, and the member component. For comparison, we also report R^2s from regressions where each trader-specific principal component is included individually.[19]

[19] As noted in the data discussion above, system order data account for only 40%-50% of NYSE volume and do not include orders handled manually by floor traders. As a result, common components in system order flow may differ from common components in total order flow. In fact, the correlation between the first principal component for TAQ trade imbalance and the first

When included individually, the first principal component from program trading explains 65.5% of the commonality in total submissions and 83.1% of the commonality in total submission imbalance. The first principal component from institutional trading explains 67.5% of commonality in total submissions and 41.1% of commonality in total submission imbalance. After controlling for program trading, the incremental R^2s for the institutional trading component are 25.6% for total submissions and 8.9% for total submission imbalance. These results suggest that both program order flow and institutional order flow play important roles in explaining the commonality in total order flow, though program trading is the dominant factor. Neither retail nor member order flow provide significant incremental explanatory power for commonality in total submission imbalance. However, member trading does provide some additional explanatory power for total submission commonality, where the incremental R^2 for member order flow is approximately 3.0%.

Given that program and institutional order flow explain the majority of order flow commonality, how much of the variation in individual security order flow can these variables explain? To address this question, we estimate a series of individual stock regressions where the dependent variable is a measure of individual security order flow and the explanatory variables are order flow principal components. The regressions take the following general form:

$$OrderFlow_{it} = \alpha_i + \sum_{type=1}^{4} \beta_{type} \cdot P_t^{type} + \varepsilon_{it}\,, \tag{6}$$

where $OrderFlow_{it}$ is firm-specific order flow and the P_t are the first principal components from program, institutional, retail, and member order flow. We focus here on total market plus limit order submission imbalance as this variable appears to have strong common components in the

principal component for total SuperDOT submission imbalance is 0.72, suggesting that the two common components are closely related but not identical. In the return analysis to follow, we include commonality in TAQ trade imbalance as an additional explanatory variable to capture commonality that may be unaccounted for in the electronic order flow data.

full sample and is closely related to the trade imbalance measures used in prior studies. Table 6 reports the average R^2s across the individual security regressions. As in Table 5, we report both individual and incremental R^2s as a measure of the explanatory power of the common components. Results are reported for the full sample and for the subsamples of large and small firms.[20]

When included individually, the program trading component explains an average of 6.5% of the variation in individual security order flow in the full sample and the institutional trading component explains an average of 3.4%. The institutional common factor provides more explanatory power for large firms than for small. This is consistent with the more active role of institutional traders among large firms. Common factors from retail and member order flow provide little explanatory power for individual security order flow even when other explanatory variables are excluded. However, the common factor in member order flow plays a more important role for large firms and the common factor in retail order flow plays a more important role for small firms.

After controlling for commonality in program order flow, the incremental R^2 for the institutional order flow common factor is approximately 1.1% and is greater for large firms than small firms. Given that the first principal component for total submission imbalance explains

[20] One possible concern is that any relation between common factors could be driven by persistence in the principal components. Like Korajczyk and Sadka (2008), we find evidence of weak autocorrelations in the principal components of return and imbalance variables, but significant autocorrelations in the principal components of volume and bid-ask spread variables. As a robustness check, we follow Korajczyk and Sadka in estimating AR(2) models for each of the principal components. We then reestimate the regressions in Tables 5, 6, 8, and 10 replacing each principal component with residuals from the associated AR(2) model. In general, our conclusions regarding the importance of program trading and the relation between returns and order flow are unchanged when principal components are replaced with their AR(2) residuals. The one notable difference in results is for Table 10, where we find that the incremental explanatory power of order flow common components in the bid-ask spread regressions is lower when AR(2) residuals are used. Overall, our main conclusions are robust to controls for persistence in the key variables of interest.

7.6% of the variation in that variable (Table 3), these results suggest that program traders account for roughly 86% of the common component and institutional traders account for the remaining 14%. Neither the retail nor member order flow common factors provide significant incremental explanatory power.

4.4. The economic interpretation of common factors

To place some economic meaning on the common factors, we examine the individual security eigenvector loadings from the first two principal components for returns and total submission imbalance.[21] We begin by plotting the eigenvector loadings for returns and order flow for the 100 sample stocks, where stocks are sorted by descending market capitalization. Loadings for returns are plotted in Panels A, B, and C of Figure 3, while loadings for order flow are plotted in Panels D, E, and F.

For both returns and order flow, the eigenvector loadings on the first principal component appear to reflect a market-wide average. This averaging is a general characteristic of the first common factor in principal components analysis. However, while large firms get more weight in the first principal component of returns, the first principal component of order flow appears to reflect equal weighting. This difference is consistent with Hasbrouck and Seppi's (2001) cautionary note that, unlike for returns, there is no theory motivating a capitalization-weighted liquidity factor. The time-series correlation between the first principal component and the market-wide equally-weighted average is 0.90 for returns and 0.98 for aggregate order flow imbalance.

What is perhaps more interesting is the interpretation of the second common factor that is

[21] A similar approach is employed by Driessen, Melenberg, and Nijman (2001) to examine common factors in international bond returns. They link factor loadings to the cross-section of bond maturities and link the factors themselves to the level, slope, and shape of the yield curve.

evident in both returns and order flow. In particular, the second principal component in each case appears to reflect differences between large and small stocks. This is consistent with the SMB return factor included in Fama-French three-factor models and with Barberis and Shleifer's (2003) style-based trading explanation for return commonality. The figures provide no clear interpretation of the third factor in either returns or order flow.

To investigate habitat and style-based explanations in more detail, we estimate cross-sectional regressions of eigenvector loadings on a set of stock characteristics. The primary variable of interest is a dummy variable to identify stocks included in the S&P 500 index during our sample period. If common factors are driven primarily by program trading or index investing, we expect eigenvector loadings to be significantly related to S&P inclusion. In addition, this variable should be most important in the order flow of program and institutional traders. To control for other factors that may affect style or habitat-based trading, we include the natural log of stock price, the natural log of market capitalization, the book-to-market ratio, and the proportion of shares held by institutions. We estimate regressions for each of the first two principal components for returns, signed TAQ trade imbalance, and signed order flow imbalance, as well as trader type specific order flow imbalance. The results are provided in Table 7.

The results for returns confirm the interpretation described above. The eigenvector loadings on the first principal component tend to be higher for large stocks, high priced stocks, value stocks, and stocks in the S&P 500 index. However, S&P inclusion appears to have only a limited effect, as the adjusted R^2 equals 75.8% excluding the S&P dummy compared to 80.5% including the S&P dummy. S&P inclusion has a more significant impact on the second principal component for returns. Here, eigenvector loadings are negatively related to both firm size and the S&P dummy, and the adjusted R^2 increases from 44.3% to 59.2% when the S&P dummy is added. These results are consistent with our interpretation of the factors loadings plotted in

Figure 3.

For signed TAQ trade imbalance, only the first principal component appears to be significantly related to firm characteristics. The eigenvector loadings from the first principal component are positively related to firm size and to S&P inclusion and the adjusted R^2 increases from 75.5% to 88.6% when the S&P dummy is added.

Turning to aggregate order flow imbalance, we see only limited explanatory power in the regressions for the first principal component loadings. The average R^2 from this regression is 10.9% and only stock price has a significant coefficient. However, the explanatory power is much higher for the second principal component. The eigenvector loadings for this principal component are positively related to firm size and S&P inclusion, with the adjusted R^2 increasing from 58.6% without the S&P dummy to 74.2% with the S&P dummy. Again, these results are consistent with our interpretation of the factor loadings plotted in Figure 3.

When order flow is broken down by trader type, the results show that firm characteristics provide very little explanatory power for eigenvector loadings in retail and member order flow. For program and institutional order flow, however, firm characteristics play a more significant role. For program traders, the importance of firm characteristics is most evident in the second principal component, where eigenvector loadings are positively related to both firm size and S&P inclusion. For institutional order flow, S&P inclusion is negatively related to eigenvector loadings on the first two principal components.

Overall, the results in this section are consistent with the role of firm characteristics and style-based explanations in determining comovements in returns and order flow. More importantly, the results support the interpretation that both returns and order flow exhibit multiple common factors driven by different economic forces and different classes of traders.

4.5 Relation between commonality in returns and order flow

It is clear from the analysis above that significant common factors exist in order flow and that these common factors are driven by program trading and institutional trading. We now examine the extent to which these order flow common factors can explain variation in returns. Specifically, we examine whether common factors in order flow have explanatory power for returns after controlling for a security's own order flow.

Hasbrouck and Seppi (2001) develop a flexible model in which the return on a stock depends on both its own order flow and the order flow of other securities. We estimate a similar model allowing the returns on security *i* to be a function of order flow on security *i* and order flow common factors. Specifically, we estimate firm-specific regressions of individual stock returns as follows:

$$R_{it} = \alpha_i + \alpha_{i0} \cdot OrderFlow_{it} + \sum_{\text{type}=1}^{4} \beta_{i,type} \cdot P_t^{type} + \beta_{i,Taq} \cdot P_t^{TAQTrades} + \varepsilon_{it}\,, \tag{7}$$

where R_{it} is the return on security *i* in period *t*, $OrderFlow_{it}$ is own order flow imbalance during period *t*, and order flow principal components are defined as in equation (6). $P_t^{TAQTrades}$ is the first principal component from signed TAQ trades and is included to account for order flow commonality that is not captured by the system order data. To examine the relative importance of order flow common components, we estimate for each stock a series of regressions where the explanatory variables are added sequentially starting with own order flow, then the program principal component, the institutional component, the retail component, the member component, and finally the TAQ trade component. We then report the cross-sectional average of the incremental R^2s from these regressions.

Results based on aggregate order submission imbalance are presented in Panel A of Table 8. In isolation, the first principal component in aggregate order submission imbalance explains

5.0% of return variation, while the common components in program and institutional order imbalances explain 4.3% and 2.6%, respectively. Even excluding all other explanatory variables, neither retail order flow nor member order flow comovements appear to provide significant explanatory power for returns. Comovements in TAQ trade imbalance explain an average of 5.6% of return variation. The ability of order flow commonality to explain return variation is higher for large stocks than for small stocks. In the large stock subsample, for example, the first principal component in aggregate order flow imbalance explains 9.4% of return variation and the first principal component in TAQ trade imbalance explains 12.7% of return variation. These results are comparable to those for the 30 Dow stocks in Hasbrouck and Seppi (2001).

We now turn to results from sequential regressions controlling for own order flow. In the full sample, own order flow accounts for 8.83% of the variation in returns.[22] The common component of aggregate order flow accounts for an additional 2.59% of return variation and the common component in TAQ trades accounts for an additional 1.34%. Thus, the magnitude of return variation explained by order flow common components is approximately half that explained by own order flow. However, the results differ substantially across firm size categories. For large firms, own order flow accounts for 6.12% of return variation, while common components in order flow and trades account for an additional 9.43%. In contrast, own order flow accounts for 10.88% of return variation in small firms, with common components adding only 0.69%. It appears that return variation for small firms is driven primarily by own order flow, while return variation for large firms is significantly affected by both own order flow

[22] Hasbrouck and Seppi (2001) find that own order flow accounts for approximately 25% of variation in returns for the 30 Dow stocks during 1994. For comparison, we repeated our analysis using signed TAQ trades as the measure of own order flow rather than signed order submissions. Using signed TAQ trades, we find that own trading accounts for 14.3% of return variation in the full sample, 18.7% in the large firm subsample, and 13.2% in the small firm subsample. While using signed TAQ trades substantially increases the explanatory power of own order flow, the incremental effects of common order flow components are similar when this variable is included.

and common order flow.

Panel B of Table 8 reports both individual and incremental R^2s based on trader-specific order flow components. After controlling for own order flow, common program order flow explains an average of 2.26% of individual security return variation in the full sample. Common institutional order flow explains an additional 0.47% and common TAQ trading explains an additional 1.24%. Neither retail nor member order flow commonality appear to provide additional explanatory power for returns. Again the results differ substantially across firm size categories. After controlling for own order flow, commonality in program order flow accounts for approximately 6.16% of return variation for large firms and 0.46% for small firms. Commonality in institutional order flow accounts for an additional 0.78% for large firms and 0.20% for small firms. Commonality in signed TAQ trades explains an additional 3.10% of return variation for large firms, but only 0.31% for small firms. We conclude that the returns of large firms are driven by own order flow, as well as commonality in program order flow, institutional order flow, and floor trading (TAQ trades). In contrast, the returns of small firms are driven almost exclusively by own order flow.[23]

[23] Following Hasbrouck and Seppi (2001), we also examine the relation between common factors in returns and order flow using canonical correlation analysis. Canonical correlations measure the strength of association between a linear combination of one set of variables and a linear combination of a second set of variables. Of the total variation in standardized returns, approximately 8% can be explained by the return canonical variable, while 6.9% can be explained by the first canonical variable related to TAQ trade imbalance and 5.4% can be explained by the first canonical variable related to order submission imbalance. More importantly, when the analysis is broken down by trader type, we find that program trading and institutional trading provide the greatest explanatory power with respect to returns. Of the total variation in returns, 4.9% can be explained by the first program trading canonical variable and 3.5% can be explained by the first institutional trading canonical variable. In contrast, retail and member order flow provide very little explanatory power for returns. Comparing results across firm size categories, we find that the program order canonical variable explains 8.3% of return variation for large firms and 2.1% of return variation for small firms. Using a lead-lag canonical correlation analysis, Korajczyk and Sadka (2008) find that shocks to monthly returns predict future liquidity, but shocks to liquidity do not predict returns. We perform a similar analysis

5. Commonality in liquidity

Several recent studies identify systematic components in measures of liquidity. For example, Chordia, Roll, and Subrahmanyam (2000) find evidence of market and industry factors in daily bid-ask spreads and quoted depth and Huberman and Halka (2001) provide evidence that time-series innovations in daily spreads and depth are correlated across stocks. Using monthly data, Korajczyk and Sadka (2008) find that a common systematic component across eight common liquidity measures accounts for most of the explained variation in individual security liquidity. Hasbrouck and Seppi (2001) also identify systematic components in intraday quoted liquidity measures, but conclude that these components are economically small. We extend these analyses in two ways. First, we test whether commonality exists in measures of liquidity beyond the quoted spread. Specifically, we estimate principal components analysis for measures of liquidity derived from limit order book depth. Second, we examine the relative importance of various types of traders in explaining variation in individual security liquidity over time.

Table 9 summarizes the principal components analysis. We present evidence for quoted bid-ask spreads and the two limit book spread measures defined in Section 2. We also provide results for limit book depth and depth imbalance at two price increments from the quote midpoint. Means and standard deviations are shown for all variables, along with the first three eigenvalues and the proportion of variance explained by the first three principal components. As in all prior analyses, principal components are estimated using standardized variables.

The full sample results (Panel A in Table 9) provide strong evidence of common components in all liquidity measures. As for returns and order flow, the results suggest that multiple common factors are present in the liquidity variables. In addition, the strength of

---

using our intraday periods and find little evidence of returns predicting future liquidity, or vice versa.

commonality increases as more information from the limit order book is incorporated. For example, the first eigenvalue for quoted spreads is 3.18 and the first three principal components explain 8.46% of the variation in this standardized variable. In contrast, the first eigenvalue for the 1,000-share spread is 5.56 and the first three principal components explain approximately 12.95% of the variation in this variable. Results based on limit book depth are similar. The first eigenvalues for depth within $0.50 of the quote midpoint and depth imbalance within $0.50 of the quote midpoint are 7.07 and 7.59, respectively.

Results for large and small firms are shown in Panels B and C in Table 9, respectively. In contrast to the findings for returns and order flow, the results in Table 9 suggest that commonality in liquidity is stronger for small firms than for large firms. The first eigenvalue for dollar quoted spreads is 1.76 for large firms and 1.95 for small firms. The first eigenvalue for depth within $0.50 of the quote midpoint is 2.27 for large firms and 3.62 for small firms. While the evidence for multiple common factors is weaker in these subsamples, there remains evidence of multiple common factors, especially for liquidity measures derived from the limit order book.

Unlike the findings for returns and order flow, the eigenvector loadings from the principal components for liquidity do not lead to simple economic interpretations with respect to firm size. For quoted spreads, the first principal component appears to place more weight on small stocks and less weight on large stocks, while the second factor places positive weights in all stocks. For the 1,000-share spread, the loadings are even more disperse. It appears that commonality in liquidity may be related to factors other than firm size. One possibility is an industry effect such as that documented by Chordia, Roll, and Subrahmanyam (2000). To conserve space, we do not plot the eigenvectors loadings for the liquidity variables.

Given the existence of common factors in liquidity, we now address three additional questions. First, to what extent is commonality in liquidity driven by commonality in order flow?

Second, are some components of order flow or some types of traders more influential in driving liquidity commonality? Third, to what extent is commonality in liquidity driven by order flow versus order type, as described in Domowitz, Hansch, and Wang (2005)? To answer these questions, we estimate firm-specific regressions of individual stock liquidity measures on trading characteristics, own order flow, own limit ratio, and common components in order flow and limit ratio, as follows:

$$
\begin{aligned}
&Liquidity_{it} = \alpha_{i0} + \alpha_{i1} \cdot AbsReturn_{it} + \alpha_{i2} \cdot TaqTrades_{it} \\
&+ \sum_{type=1}^{4} \left(\beta_{i,type} \cdot OrderFlow_{it}^{type} + \delta_{i,type} \cdot LmtRatio_{it}^{type} + \gamma_{i,type} \cdot PFLOW_{t}^{type} + \lambda_{i,type} \cdot PLMT_{t}^{type} \right) + \varepsilon_{it},
\end{aligned} \tag{8}
$$

where $Liquidity_{it}$ is the liquidity measure for security i in period t, $AbsReturn_{it}$ is the absolute midpoint return on security i during period t, $TaqTrades_{it}$ is the number of TAQ trades for security i in period t, $OrderFlow_{it}$ is own order flow for a given trader type during period t, and $LmtRatio_{it}$ is the ratio of limit order submissions to total submissions for a given trader type during period t. Order flow principal components ($PFLOW^{type}$) are defined for the four trader types as in equation (6) and limit order ratio principal components ($PLMT^{type}$) are defined analogously for *LmtRatio*. The order flow variable analyzed here is total market plus limit order submissions. To examine the incremental explanatory power of trader-specific order flow and trader-specific common components, we estimate for each stock a series of regressions where the explanatory variables are added sequentially starting with own absolute returns and TAQ trades, then trader-specific order flow, trader-specific order flow principal components, trader-specific limit ratio, and trader-specific limit ratio principal components.

Table 10 reports the average incremental R^2s from the security-specific regressions, with results for the full sample reported in Panel A and results for the large and small stock subsamples reported in Panels B and C, respectively. Results are reported for the quoted spread, the limit book spread, and the 1,000-share spread. For the full sample, absolute returns and TAQ trades explain approximately 4.98% of the variation in quoted bid-ask spreads. Adding own order flow by trader type explains an additional 0.85%, and adding common order flow components explains an additional 1.51%. The incremental explanatory power of order flow common components is similar for the limit spread and 1,000-share spread, though the overall level of explanatory power for these dependent variables is lower.

Comparing results across firm size categories, we see that the incremental explanatory power of common order flow components is substantially higher for small firms than for large firms. For example, the common order flow components provide an incremental R^2 of 0.92% for quoted spreads in the large firm subsample and 2.05% for quoted spreads in the small firm subsample. The incremental explanatory power of own order flow is also higher for small firms than for large firms. Together, these results suggest that individual security liquidity measures are significantly influenced by both trader-specific order flow and by common components in order flow. In contrast to the return results, the relation between liquidity and order flow commonality is stronger for small firms than for large firms.[24]

Consistent with Domowitz, Hansch, and Wang (2005), the results suggest that both own limit order use and commonality in limit order use have incremental explanatory power for liquidity. For example, adding own limit ratio to the regressions for quoted spreads increases the R^2 from 7.24% to 7.99% and adding the first principal component for limit ratio increases the R^2 to 8.73%. The explanatory power related to limit order use is higher for small securities than for large securities. For example, adding both own limit ratio and the first principal component for limit ratio increases the R^2 from 6.49% to 7.38% for large stocks and from 6.59% to 8.57% for

[24] As noted earlier, the incremental explanatory power of the common order flow components in these liquidity regressions is reduced if we replace each principal component with its associated residuals from an AR(2) model. However, the incremental explanatory power of common order flow components continues to be higher for small firms than for large firms and the explanatory power of the limit order ratio principal component is similar to that reported above.

small securities.

While the analysis in Table 10 highlights the relative importance of various order flow components in explaining liquidity, it does not identify the direction of these effects. As noted in Chordia, Roll, and Subrahmanyam (2000), order flow can have either positive or negative effects on execution costs depending on whether the order flow reflects liquidity trading or information trading. They also note that broad market activity is more likely to affect inventory risk, while individual trading activity is more likely related to informed trading. To analyze these effects by trader type, we examine the cross-sectional averages of coefficients from the security-specific regressions described in equation (8). These results are described below and are available from the authors.

After controlling for absolute returns and TAQ trades, increases in both own order submissions and common order submissions result in decreased execution costs. This could reflect either a decrease in the inventory risk of the market-maker or an increase in liquidity (limit orders) provided by market participants other than the specialist. Both own limit order use and commonality in limit order use have significant negative effects on liquidity. These findings are consistent with Domowitz, Hansch, and Wang (2005) and suggest that comovements in supply and demand are a significant determinant of liquidity commonality. The results for trader specific order flow show that the negative effect of own submissions is driven by program and institutional traders, while the negative effect of common order flow submissions is driven primarily by non-program traders. In addition, the coefficient on own limit ratio appears to be significantly negative for program, institutional, and retail traders, while the negative coefficient on common limit order ratio is driven by program traders and institutional traders only.

Consistent with Chordia, Roll, and Subrahmanyam (2000) and Huberman and Halka (2001), our results suggest that common factors exist in liquidity measures for NYSE-listed stocks. While common factors explain only a small fraction of variation in individual security quoted spreads, these factors are more important for measures of liquidity based on limit order book depth. We find that individual security liquidity measures are significantly related to both own order flow and common factors in order flow and that these effects are strongest for the smallest stocks. In addition, our results are consistent with Domowitz, Hansch, and Wang (2005), who show that commonality in liquidity is driven by comovements in order type.

6. Conclusion

Prior research identifies significant common factors in returns, order flow, and liquidity. However, this research provides only limited evidence as to the sources of commonality. Using a unique dataset consisting of all electronic order flow for a sample of NYSE stocks, we examine commonality by trader type and test the relative importance of trader-specific factors in explaining commonality in order flow, returns, and liquidity.

We show that common factors are evident in numerous measures of order flow based on both submissions and executions of market and limit orders. Common factors explain approximately 8% of the variation in signed market plus limit order submissions. Program trading accounts for nearly all of the commonality in order flow, with other institutional trading playing a lesser role. Though common factors appear to exist in both retail and member order flow, neither of these appears to significantly affect overall order flow commonality.

We find that commonality in order flow is significantly related to returns. In particular, both commonality in program order flow and commonality in institutional order flow provide incremental explanatory power for returns on large firms, after controlling for own order flow. In contrast, small firm returns are driven primarily by own order flow and are only marginally related to order flow common factors. Notably, we find that contrasts between small and large firms play an important role in both return and order flow common factors and that loadings on common factors are significantly related to firm characteristics. These findings are consistent with style-based explanations of commonality, as described in Barberis and Shleifer (2003).

Consistent with prior research, we find evidence of weak common factors in quoted bid-ask spreads. Commonality is stronger for measures of liquidity based on limit order book depth and is stronger for small firms than for large firms. After controlling for number of trades and absolute returns, we find that common order flow components tend to be negatively related to individual security spreads, regardless of trader type. Consistent with Domowitz, Hansch, and Wang (2005), we find that commonality in liquidity is driven in part by comovements in order type.

Our findings are consistent with style and habitat-based explanations of commonality in returns and order flow. In addition, we show that commonality is driven in large part by order flow from program traders. These findings suggest that the importance of commonality may be increasing over time, as program trading becomes a larger and larger fraction of overall trading volume. Consistent with this conjecture, Kamara, Lou, and Sadka (2008) find that the sensitivity to market liquidity has increased over time for large stocks and decreased over time for small stocks. The effects we document have important implications for our understanding of portfolio theory, trading decisions, and the relations between trading activity, prices, and liquidity.

References

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

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

Bloomfield, R., O'Hara, M., Saar G., 2005. The 'make or take' decision in an electronic market: Evidence on the evolution of liquidity. Journal of Financial Economics 75, 165–199.

Bodurtha, J., Kim, D., Lee, C. M., 1995. Closed-end country funds and U.S. market sentiment. Review of Financial Studies 8, 879–918.

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

Chan, L., Lakonishok, J., 1995. The behavior of stock prices around institutional trades. Journal of Finance 50, 1147–1174.

Chan, L., Lakonishok, J., 1997. Institutional equity trading costs: NYSE versus NASDAQ. Journal of Finance 70, 713–735.

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

Coughenour, J. F., Saad, M. M., 2004. Common market makers and commonality in liquidity. Journal of Financial Economics 73, 37–70.

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

Driessen, J., Melenberg, B., Nijman, T., 2003. Common factors in international bond returns. Journal of International Money and Finance 22, 629–656.

Edelen, R. M., Warner, J. B., 2001. Aggregate price effects of institutional trading: A study of mutual fund flow and market returns. Journal of Financial Economics 59, 195–220.

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

Fama, E., French, K., 1995. Size and book-to-market factors in earnings and returns. Journal of Finance 50, 131–155.

Foster, F. D., Viswanathan, S., 1993. Variation in trading volume, return volatility, and trading costs: Evidence on recent price formation models. Journal of Finance 48, 187–211.

Froot, K., Dabora, E., 1999. How are stock prices affected by the location of trade? Journal of Financial Economics 53, 189–216.

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

Harris, L., Sofianos, G., Shapiro, J. E., 1994. Program trading and intraday volatility. Review of Financial Studies 7, 653–685.

Hasbrouck, J., 1992. Using the TORQ database. Working paper, New York Stock Exchange.

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

Huberman, G., Halka, D., 2001. Systematic liquidity. Journal of Financial Research 24, 161–178.

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

Johnson, R. A., Wichern, D. W., 2007. Applied Multivariate Statistical Analysis. 6th edition, Prentice Hall.

Jones, C., Lipson, M., 1999. Execution costs of institutional equity orders. Journal of Financial Intermediation 8, 123–140.

Kamara, A., Lou, X., Sadka, R., 2008. The divergence of liquidity commonality in the cross-section of stocks. Journal of Financial Economics 89, 444–466.

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

Kavajecz, K., 1999. The specialist's quoted depth and the limit order book. Journal of Finance 54, 747–771.

Keim, D., Madhavan, A., 1997. Transactions costs and investment style: An inter-exchange comparison of equity trades. Journal of Financial Economics 46, 265–292.

Keim, D., Madhavan, A., 1996. The upstairs market for block trades: Analysis and measurement of price effects. Review of Financial Studies 9, 1–36.

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

Kumar, A., Lee, C., 2006. Retail investor sentiment and return comovements. Journal of Finance 61, 2451–2486.

Lee, C., Shleifer, A., Thaler, R., 1991. Investor sentiment and the closed-end fund puzzle. Journal of Finance 46, 75–110.

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

McInish, T. H., Wood, R. A., 1992. An analysis of intraday patterns in bid-ask spreads for NYSE stocks. Journal of Finance 47, 753–764.

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

New York Stock Exchange, 1998. Fact Book. New York Stock Exchange, New York.

Sofianos, G., Werner, I. M., 2000. The trades of NYSE floor brokers. Journal of Financial Markets 3, 139–176.

Vijh, A., 1994. S&P trading strategies and stock betas. Review of Financial Studies 7, 215–251.

Warther, V., 1995. Aggregate mutual fund flows and security returns. Journal of Financial Economics 39, 209–235.

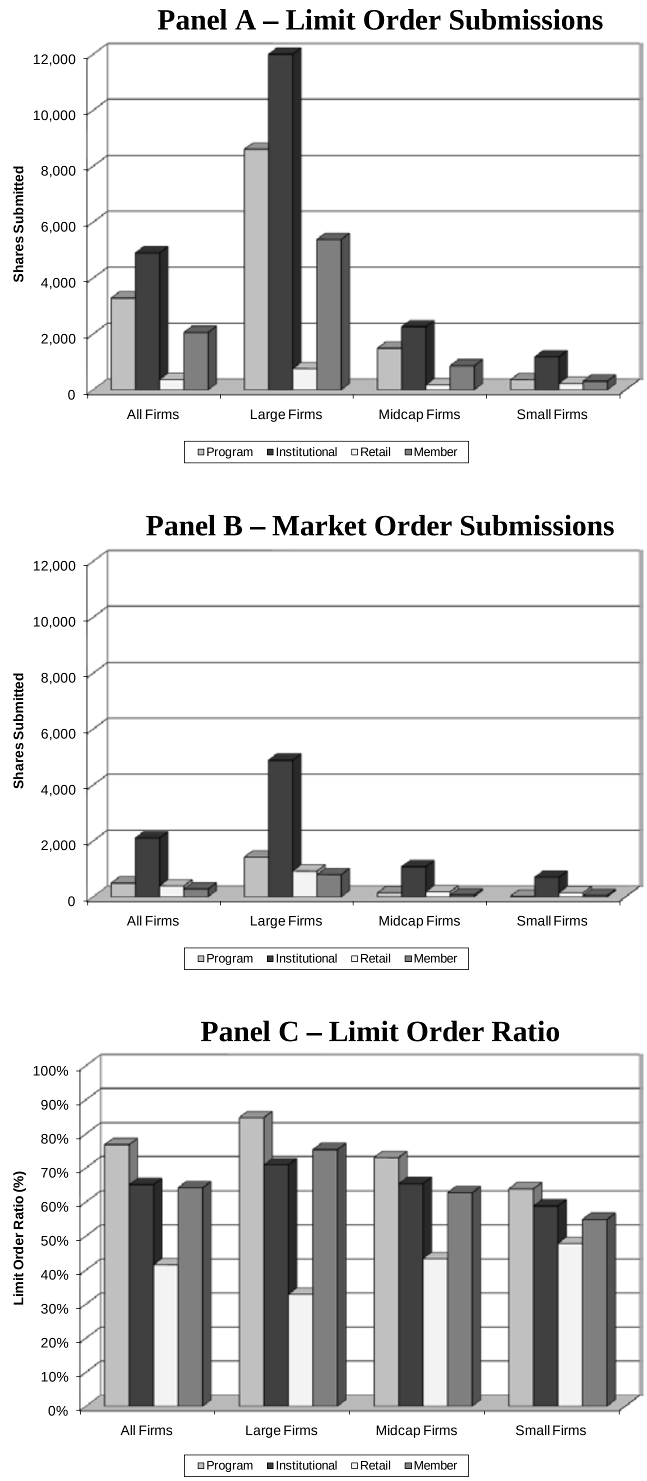
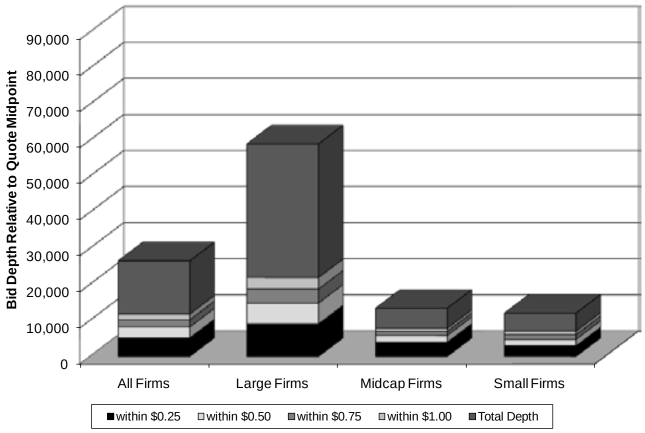


Figure 1. Order submissions and limit order use by trader type and firm size

The figure plots average limit and market order submissions and the proportion of submitted shares that are limit orders per 15 minute period. Trader categories include program traders, institutional traders, retail traders, and exchange members. Firms are categorized by market capitalization, where small, midcap, and large securities are defined as those in market capitalization deciles 1-3, 4-7, and 8-10, respectively. The data include all electronic orders submitted through the SuperDOT system from November 1997 through February 1998. The sample includes 100 NYSE-listed securities selected randomly across deciles of market capitalization.



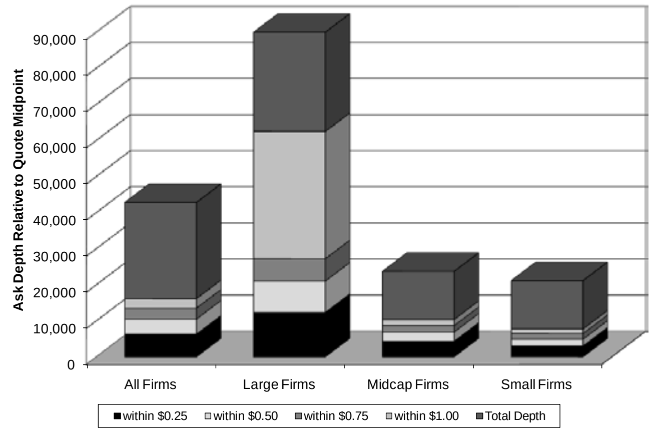


Figure 2. Limit order book characteristics by firm size

The figure illustrates cumulative limited order book depth by firm size category. Firms are categorized by market capitalization, where small, midcap, and large securities are defined as those in market capitalization deciles 1-3, 4-7, and 8-10, respectively. Limit order books are reconstructed using the methodology of Kavajecz (1999). The data include all electronic orders submitted through the SuperDOT system from November 1997 through February 1998. The sample includes 100 NYSE-listed securities selected randomly across deciles of market capitalization.

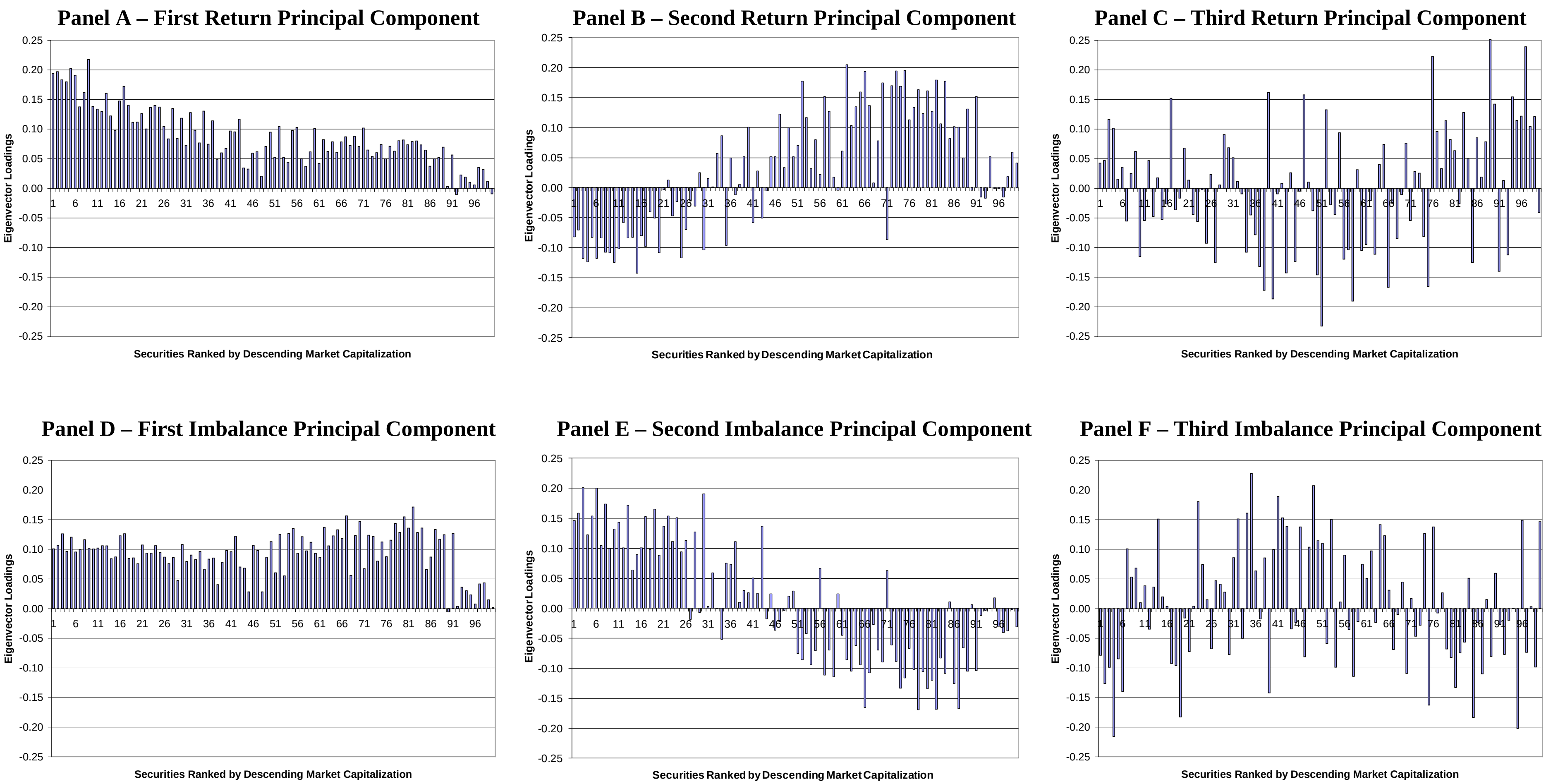


Figure 3. Eigenvectors from principal components analysis on returns and total submission imbalance

The figure plots the eigenvector loadings from the first three principal components for returns (Panels A, B, and C) and total submission imbalance (Panels D, E, and F). Returns are defined as the natural log of the ratio of ending quote midpoint and beginning quote midpoint from each 15-minute interval. Total submission imbalance is defined as the share volume of market and limit buy orders minus the share volume of market and limit sell orders divided by total shares submitted. The data include all electronic orders submitted through the SuperDOT system from November 1997 through February 1998. The sample includes 100 NYSE-listed securities selected randomly across deciles of market capitalization.

Table 1. Sample characteristics

The table lists means (medians) for each variable. The sample includes 100 NYSE-listed securities selected randomly across market capitalization deciles and the sample period is from November 1997 through February 1998. Panel A describes firm characteristics. Trade data are from TAQ. Price and Market Capitalization are as of October 31, 1997 and are from CRSP. Institutional holdings and book-to-market ratio are calculated as of year-end 1997. Book-to-market is defined as in Fama and French (1993). Panel B describes quoted bid-ask spreads and three measures of spreads estimated based on limit book depth, where the limit order book is reconstructed using the methodology of Kavajecz (1999). For each firm, spreads are averaged across all 15-minute intervals during the sample period. Results are provided for the full sample and for subsamples categorized by market capitalization, where small, midcap, and large securities are defined as those in market capitalization deciles 1-3, 4-7, and 8- 10, respectively. The *p*-value is from a test of the restriction that means (medians) are equal across size categories.

	All Securities (N=100)	Market Capitalization Category: Large (N=30)	Midcap (N=40)	Small (N=30)	*p*-value
Panel A: Firm Characteristics					
Market Capitalization ($ mil)	5566.88	16529.24	1292.61	303.55	0.000
	(1122.64)	(7500.84)	(1122.64)	(297.37)	(0.000)
Price ($)	37.80	54.96	36.65	22.16	0.000
	(31.04)	(52.08)	(31.04)	(22.85)	(0.000)
Daily Share Volume (000)	316.68	814.15	135.69	60.54	0.000
	(98.40)	(582.56)	(85.52)	(27.82)	(0.000)
Daily Number of Trades	203.03	512.90	81.43	55.30	0.000
	(76.00)	(282.50)	(64.50)	(24.00)	(0.000)
Std. Dev. of Daily Returns (%)	2.06	1.88	2.10	2.19	0.509
	(1.87)	(1.87)	(1.74)	(1.90)	(0.576)
Book-to-Market Ratio	0.49	0.62	0.49	0.35	0.000
	(0.42)	(0.55)	(0.48)	(0.30)	(0.001)
Institutional Shareholdings (%)	55.61	44.29	57.51	64.38	0.001
	(59.00)	(46.36)	(62.29)	(63.38)	(0.003)
Panel B: Bid-Ask Spreads					
Quoted Spread (bps)	62.59	29.62	58.60	100.88	0.000
	(53.67)	(29.50)	(56.26)	(88.38)	(0.000)
Limit Spread (bps)	203.99	40.87	186.49	390.44	0.000
	(137.26)	(36.89)	(139.43)	(316.74)	(0.000)
1000-Share Spread (bps)	286.81	75.36	262.22	531.04	0.000
	(210.69)	(76.60)	(221.48)	(466.13)	(0.000)

Table 2. Principal components analysis for return and trading variables

For each order flow and trading variable, the table lists the first three eigenvalues from the principal components analysis along with the proportion of total variance explained by these three eigenvalues. Under the assumption of no commonality, the cumulative explained variance would equal 3/N. All variables are calculated at 15-minute intervals from 9:30 a.m. through 4:00 p.m. In addition, all variables used in the principal components analysis are standardized using firm and time-of-day specific means and standard deviations. Trade data are from TAQ and are signed using the methodology of Lee and Ready. Returns are defined based on quote midpoints at the beginning and end of each period. Imbalances are defined as buy volume minus sell volume divided by total volume. Means and standard deviations are for the raw variables, pooled across firms and time. For these statistics, volume is reported in thousands of shares, imbalances are reported as a percentage of total share volume, and returns are reported as percentages. The sample includes 100 NYSE-listed securities selected randomly across market capitalization deciles and the sample period is from November 1997 through February 1998. Results for the full sample are provided in Panel A. Panels B and C provide results for subsamples categorized by market capitalization, where small and large stocks are defined as deciles 1-3 and 8-10, respectively.

	Mean	Std. Dev.	Eigenvalues			Cumulative Explained Variance (%)
			1	2	3	
Panel A: All Securities (*N*=100)						
Return	0.00	0.53	8.43	1.82	1.42	11.65
Absolute Return	0.19	0.49	4.91	1.73	1.56	8.18
# of Trades (All)	7.77	17.40	8.14	3.08	2.14	13.36
# of Trades (Large)	0.21	0.80	1.75	1.58	1.47	4.09
# of Trades (Small)	3.82	9.82	6.21	2.70	2.13	11.04
Trade Imbalance (All)	1.04	58.41	4.90	1.56	1.42	7.87
Trade Imbalance (Large)	0.95	32.08	1.39	1.35	1.35	5.07
Trade Imbalance (Small)	0.99	58.18	3.04	1.66	1.54	6.23
Panel B: Large Securities (*N*=30)						
Return	0.00	0.33	5.96	1.11	1.05	27.05
Absolute Return	0.21	0.26	3.60	1.19	1.12	19.70
# of Trades (All)	19.67	26.93	5.99	1.74	1.42	30.49
# of Trades (Large)	0.57	1.32	1.53	1.35	1.27	13.92
# of Trades (Small)	9.91	15.78	4.61	1.73	1.51	26.16
Trade Imbalance (All)	3.85	49.03	3.99	1.18	1.15	21.05
Trade Imbalance (Large)	2.29	48.75	2.50	1.29	1.29	16.91
Trade Imbalance (Small)	3.34	58.11	1.24	1.23	1.18	12.25
Panel C: Small Securities (*N*=30)						
Return	0.00	0.41	2.01	1.20	1.19	14.58
Absolute Return	0.17	0.37	1.59	1.32	1.23	13.77
# of Trades (All)	2.09	7.53	1.93	1.58	1.24	15.83
# of Trades (Large)	1.00	3.72	1.08	1.02	1.01	10.36
# of Trades (Small)	0.03	0.20	1.73	1.55	1.21	14.96
Trade Imbalance (All)	-0.92	57.18	1.28	1.21	1.17	12.21
Trade Imbalance (Large)	-0.35	50.44	1.12	1.04	0.97	10.43
Trade Imbalance (Small)	0.08	14.39	1.22	1.20	1.18	12.00

Table 3. Principal components analysis for order flow variables

For each order flow variable, the table lists the first three eigenvalues from the principal components analysis along with the proportion of total variance explained by these three eigenvalues. Under the assumption of no commonality, the cumulative explained variance would equal 3/N. All variables are calculated at 15-minute intervals from 9:30 a.m. through 4:00 p.m. In addition, all variables used in the principal components analysis are standardized using firm and time-of-day specific means and standard deviations. Order flow data include all orders submitted through the NYSE's electronic SuperDOT system. Imbalances are defined as buy volume minus sell volume divided by total volume. Limit order use is the proportion of submitted shares that are from limit orders. Means and standard deviations are for the raw variables, pooled across firms and time. For these statistics, submissions and executions are reported in thousands of shares, imbalances and limit order use are reported as a percentage of total shares, and returns are reported as percentages. The sample includes 100 NYSE-listed securities selected randomly across market capitalization deciles and the sample period is from November 1997 through February 1998. Results for the full sample are provided in Panel A. Panels B and C provide results for subsamples categorized by market capitalization, where small and large stocks are defined as deciles 1-3 and 8-10, respectively.

	Mean	Std. Dev.	Eigenvalues			Cumulative Explained Variance (%)
			1	2	3	
Panel A: All Securities (*N*=100)						
Total Subm (000)	14.80	38.36	5.40	2.65	1.87	9.92
Limit Order Subm (000)	11.72	26.33	5.21	2.77	1.85	9.82
Market Order Subm (000)	3.09	21.64	4.16	1.73	1.52	7.40
Total Submission Imbal (%)	5.84	64.22	7.61	2.84	1.54	11.98
Limit Submission Imbal (%)	7.56	65.14	8.27	3.00	1.56	12.82
Market Submission Imbal (%)	-0.20	64.12	5.23	1.68	1.66	8.57
Limit Order Use (%)	74.77	27.69	5.81	2.73	1.54	10.08
Panel B: Large Securities (*N*=30)						
Total Subm (000)	37.09	55.76	4.11	1.44	1.31	22.85
Limit Order Subm (000)	29.68	40.40	3.96	1.46	1.33	22.46
Market Order Subm (000)	7.41	26.78	3.06	1.26	1.21	18.43
Total Submission Imbal (%)	13.06	51.79	3.92	1.22	1.14	20.96
Limit Submission Imbal (%)	14.28	56.57	3.62	1.21	1.20	20.10
Market Submission Imbal (%)	4.44	69.82	4.29	1.22	1.12	22.09
Limit Order Use (%)	81.35	18.36	3.85	1.12	1.09	20.19
Panel C: Small Securities (*N*=30)						
Total Subm (000)	3.28	20.69	1.80	1.29	1.23	14.37
Limit Order Subm (000)	2.34	8.70	1.86	1.26	1.23	14.52
Market Order Subm (000)	0.94	18.21	1.40	1.30	1.23	13.11
Total Submission Imbal (%)	-0.81	67.20	3.36	1.19	1.17	19.04
Limit Submission Imbal (%)	1.64	64.66	4.00	1.24	1.18	21.39
Market Submission Imbal (%)	-2.87	51.84	1.27	1.25	1.18	12.46
Limit Order Use (%)	67.02	31.07	2.45	1.29	1.21	16.50

Table 4. Principal Components Analysis for Order Flow Variables by Trader Type

For each order flow variable, the table lists the proportion of total variance explained by the first eigenvalue in the principal components analysis. Under the assumption of no commonality, the explained variance would equal 1/*N*. Principal components are estimated for total order flow and for order flow categorized by trader type, where trader types include program traders, institutional traders, retail traders, and exchange members. All variables are calculated at 15-minute intervals from 9:30 a.m. through 4:00 p.m. In addition, all variables used in the principal components analysis are standardized using firm and time-of-day specific means and standard deviations. Order flow data include all orders submitted through the NYSE's electronic SuperDOT system. Imbalances are defined as buy volume minus sell volume divided by total volume. Limit order use is the proportion of submitted shares that are from limit orders. The sample includes 100 NYSE-listed securities selected randomly across market capitalization deciles and the sample period is from November 1997 through February 1998. Results for the full sample are provided in Panel A. Panels B and C provide results for subsamples categorized by market capitalization, where small and large stocks are defined as deciles 1-3 and 8-10, respectively.

		Trader Type			
	All Order Flow	Program Trading	Institutional Trading	Retail Trading	Member Trading
Panel A: All Securities (*N*=100)					
Total Subm (000)	5.40	11.01	2.96	2.29	2.09
Limit Order Subm (000)	5.21	10.45	3.05	1.76	2.04
Market Order Subm (000)	4.16	9.15	1.94	2.24	1.57
Total Submission Imbal (%)	7.61	17.66	3.05	1.85	1.71
Limit Submission Imbal (%)	8.27	18.07	3.03	1.53	1.71
Market Submission Imbal (%)	5.23	16.67	2.27	1.93	1.68
Limit Order Use (%)	5.81	11.25	2.74	1.62	2.38
Panel B: Large Securities (*N*=30)					
Total Subm (000)	13.70	27.87	7.13	5.03	5.80
Limit Order Subm (000)	13.20	25.17	7.40	4.50	5.53
Market Order Subm (000)	10.20	28.73	5.30	5.20	4.93
Total Submission Imbal (%)	13.07	26.23	6.53	4.97	4.57
Limit Submission Imbal (%)	12.07	24.67	6.30	4.37	4.43
Market Submission Imbal (%)	14.30	45.13	6.03	4.97	5.37
Limit Order Use (%)	12.83	23.90	6.00	4.43	5.87
Panel C: Small Securities (*N*=30)					
Total Subm (000)	6.00	12.77	5.20	4.43	4.03
Limit Order Subm (000)	6.20	13.27	5.53	4.20	4.03
Market Order Subm (000)	4.67	6.33	4.13	4.47	3.53
Total Submission Imbal (%)	11.20	24.93	5.40	4.13	3.97
Limit Submission Imbal (%)	13.33	26.03	6.00	4.10	3.97
Market Submission Imbal (%)	4.23	5.87	4.27	4.20	3.47
Limit Order Use (%)	8.17	19.20	5.93	4.07	4.10

Table 5. Drivers of order flow principal components

The table describes regressions of total order flow principal components on principal components based on trader type. The dependent variable is the first eigenvalue from a principal components analysis on total submissions (or total submission imbalance). The explanatory variables are the first eigenvalues from principal components analyses o trader-specific submissions (submission imbalances). All variables are calculated at 15-minute intervals from 9:30 a.m. through 4:00 p.m. In addition, all variables used in the principal components analysis are standardized using firm and time-of-day specific means and standard deviations. Order flow data include all orders submitted through the NYSE's electronic SuperDOT system. Imbalances are defined as buy volume minus sell volume divided by total volume. The sample includes 100 NYSE-listed securities selected randomly across market capitalization deciles and the sample period is from November 1997 through February 1998.

Trader Specific Principal Component	First Principal Components from Total Submissions			First Principal Components from Total Submission Imbalance		
	Individual R^2	Cumulative R^2	Incremental R^2	Individual R^2	Cumulative R^2	Incremental R^2
Program Traders	0.6545	0.6545	0.6545	0.8307	0.8307	0.8307
Institutional Traders	0.6747	0.9108	0.2563	0.4107	0.9199	0.0892
Retail Traders	0.1103	0.9108	0.0000	0.0000	0.9206	0.0007
Member Traders	0.4435	0.9407	0.0299	0.0057	0.9211	0.0005

Table 6. Explained order flow variation

The table lists the average R-square from stock-specific regressions of individual order submission imbalance on various order flow principal components. All variables are calculated at 15-minute intervals from 9:30 a.m. through 4:00 p.m. In addition, all variables used in the principal components analysis are standardized using firm and time-of-day specific means and standard deviations. The order flow data include all orders submitted through the NYSE's electronic SuperDOT system. Order submission imbalances are defined as market plus limit buy orders minus market plus limit sell orders divided by total submissions. The sample includes 100 NYSE-listed securities selected randomly across market capitalization deciles and the sample period is from November 1997 through February 1998. Results are provided for the full sample and for subsamples categorized by market capitalization, where small and large stocks are defined as deciles 1-3 and 8-10, respectively.

	All Securities (*N*=100)			Large Securities (*N*=30)			Small Securities (*N*=30)		
	Individual R^2	Cumulative R^2	Incremental R^2	Individual R^2	Cumulative R^2	Incremental R^2	Individual R^2	Cumulative R^2	Incremental R^2
Common Program Order Subm Imb	0.0653	0.0653	0.0653	0.0948	0.0948	0.0948	0.0858	0.0858	0.0858
Common Institutional Order Subm Imb	0.0343	0.0758	0.0105	0.0584	0.1152	0.0204	0.0188	0.1007	0.0149
Common Retail Order Subm Imb	0.0023	0.0781	0.0023	0.0024	0.1178	0.0026	0.0070	0.1077	0.0070
Common Member Order Subm Imb	0.0031	0.0807	0.0026	0.0122	0.1223	0.0045	0.0015	0.1093	0.0016

Table 7. Security characteristics and principal component weights

For each return and order flow principal component, we regress the eigenvectors from the principal components analysis on firm characteristics. All variables used in the principal components analysis are calculated at 15-minute intervals from 9:30 a.m. through 4:00 p.m. and are standardized using firm and time-of-day specific means and standard deviations. Returns are defined based on quote midpoints at the beginning and end of each period. The order flow data include all orders submitted through the NYSE's electronic SuperDOT system. Order submission imbalances are defined as market plus limit buy orders minus market plus limit sell orders divided by total submissions. Trade imbalance is defined using all TAQ executions as buy trades minus sell trades divided by total trades, where trades are signed using the methodology of Lee and Ready. The explanatory variables include the natural log of stock price, the natural log of market value, book-to-market ratio, fraction of shares held by institutions, and a dummy variable to identify firms included in the S&P 500 index. For comparison, the table provides the adjusted R^2 from regressions including and excluding the S&P 500 dummy variable. The sample includes 100 NYSE-listed securities selected randomly across market capitalization deciles and the sample period is from November 1997 through February 1998. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Principal Component	Explanatory Variables						Adj. R^2 excluding S&P Dummy	Adj. R^2 including S&P Dummy
		Intercept	Ln(Prc)	Ln(MV)	Bk/Mkt	Inst%	S&P Dummy		
Return	1	-0.216***	0.011**	0.017***	0.018**	0.008	0.035***	0.7576	0.8054
Return	2	0.267***	0.021	-0.019**	-0.006	0.017	-0.121***	0.4434	0.5923
TAQ Trade Imbalance	1	-0.197***	0.005	0.016***	0.007	0.011	0.071***	0.7545	0.8862
TAQ Trade Imbalance	2	0.123	0.007	-0.007	-0.064	0.011	-0.049	0.0247	0.0378
Total Submission Imbalance	1	-0.012	0.026***	0.001	0.004	0.014	-0.012	0.1082	0.1090
Total Submission Imbalance	2	-0.347***	-0.021	0.026***	0.031	0.004	0.128***	0.5863	0.7417
Program Trader Imbalance	1	0.008	0.010	0.004	0.003	0.018	-0.033***	-0.0101	0.0628
Program Trader Imbalance	2	-0.285***	0.000	0.017***	0.005	0.001	0.154***	0.6720	0.9017
Institutional Trader Imbalance	1	-0.094	0.024**	0.005	0.009	0.066***	-0.027*	0.1599	0.1846
Institutional Trader Imbalance	2	0.152	0.058**	-0.023*	-0.014	0.017	-0.061**	0.1436	0.1710
Retail Trader Submission Imbalance	1	0.164	-0.036	0.000	-0.033	-0.055	0.051	0.0000	0.0140
Retail Trader Submission Imbalance	2	-0.283**	0.014	0.019*	0.057*	-0.037	0.009	0.1170	0.1084
Member Submission Imbalance	1	-0.062	-0.004	0.007	0.010	-0.051	0.057*	0.0421	0.0652
Member Submission Imbalance	2	-0.117	-0.040	0.019	0.037	-0.028	-0.019	-0.0010	-0.0082

Table 8. Explained return variation

The table lists the average R-square from stock-specific regressions of returns on stock-specific order flow variables and order flow principal components. All variables are calculated at 15-minute intervals from 9:30 a.m. through 4:00 p.m. In addition, all variables used in the principal components analysis are standardized using firm and time-of-day specific means and standard deviations. Returns are defined based on quote midpoints at the beginning and end of each period. The order flow data include all orders submitted through the NYSE's electronic SuperDOT system. Order submission imbalances are defined as market plus limit buy orders minus market plus limit sell orders divided by total submissions. Trade imbalance is defined using all TAQ executions as buy trades minus sell trades divided by total trades, where trades are signed using the methodology of Lee and Ready. Panel A provides describes regressions in which common components are included individually. Panels B and C describe results from sequential regressions controlling for own order flow. The sample includes 100 NYSE-listed securities selected randomly across market capitalization deciles and the sample period is from November 1997 through February 1998. Results are provided for the full sample and for subsamples categorized by market capitalization, where small and large stocks are defined as deciles 1-3 and 8-10, respectively.

	All Securities (*N*=100)			Large Securities (*N*=30)			Small Securities (*N*=30)		
	Individual R^2	Cumulative R^2	Incremental R^2	Individual R^2	Cumulative R^2	Incremental R^2	Individual R^2	Cumulative R^2	Incremental R^2
Panel A (Total Order Submission Imbalance):									
Own Order Subm Imb	0.0883	0.0883	0.0883	0.0612	0.0612	0.0612	0.1088	0.1088	0.1088
Common Order Subm Imb	0.0495	0.1142	0.0259	0.0943	0.1246	0.0634	0.0215	0.1133	0.0045
Common Trade Imbalance	0.0564	0.1276	0.0134	0.1266	0.1615	0.0309	0.0154	0.1161	0.0024
Panel B (Order Submission Imbalance by Trader Type):									
Own Order Subm Imb	0.0883	0.0883	0.0883	0.0612	0.0612	0.0612	0.1088	0.1088	0.1088
Common Program Order Subm Imb	0.0429	0.1109	0.0226	0.0788	0.1228	0.0616	0.0201	0.1134	0.0046
Common Institutional Order Subm Imb	0.0261	0.1156	0.0047	0.0504	0.1306	0.0078	0.0088	0.1154	0.0020
Common Retail Order Subm Imb	0.0007	0.1164	0.0008	0.0007	0.1324	0.0018	0.0006	0.1166	0.0012
Common Member Order Subm Imb	0.0009	0.1176	0.0012	0.0009	0.1342	0.0018	0.0013	0.1184	0.0018
Common Trade Imbalance	0.0564	0.1300	0.0124	0.1266	0.1652	0.0310	0.0154	0.1215	0.0031

Table 9. Principal components analysis for limit book and execution cost variables

For each limit book and execution cost variable, the table lists the first three eigenvalues from the principal components analysis along with the proportion of total variance explained by these three eigenvalues. Under the assumption of no commonality, the cumulative explained variance would equal 3/*N*. All variables are calculated at 15-minute intervals from 9:30 a.m. through 4:00 p.m. In addition, all variables used in the principal components analysis are standardized using firm and time-of-day specific means and standard deviations. Order flow data include all orders submitted through the NYSE's electronic SuperDOT system and limit order books are reconstructed using the methodology of Kavajecz (1999). Execution costs measures include the quoted bid-ask spread and three measures of bid-ask spread calculated based on limit order book depth. The limit spread is defined as the difference between the best ask and bid prices in the limit order book. The 1,000-share spread is defined as the difference between the 1,000-share ask and the 1,000-share bid, where the 1,000-share ask (bid) is the price one would have to go up (down) to in order to reach 1,000-share of cumulative depth. 5,000-share spreads are defined similarly. Limit book measures are defined based on the cumulative share depth (depth imbalance) within $0.50 and $1.00 of the quote midpoint. Means and standard deviations are for the raw variables, pooled across firms and time. For these statistics, depth is reported in thousands of shares and imbalances are reported as a percentage of total share depth. The sample includes 100 NYSE-listed securities selected randomly across market capitalization deciles and the sample period is from November 1997 through February 1998. Results for the full sample are provided in Panel A. Panels B and C provide results for subsamples categorized by market capitalization, where small and large stocks are defined as deciles 1-3 and 8-10, respectively.

	Mean	Std. Dev.	Eigenvalues 1	Eigenvalues 2	Eigenvalues 3	Cumulative Explained Variance (%)
Panel A: All Securities (*N*=100)						
Bid-Ask Spread:						
Quoted Spread (%)	0.62	0.54	3.18	3.01	2.28	8.46
Limit Spread (%)	1.50	2.56	4.24	2.97	2.68	9.88
1,000-share Spread (%)	2.48	3.45	5.56	4.28	3.11	12.95
Limit Book Depth:						
Depth within $0.50 (000)	18.90	27.18	7.07	4.53	3.99	15.55
Depth within $1.00 (000)	28.24	38.95	9.52	5.72	5.16	20.34
Depth Imbalance within $0.50 (%)	-6.16	65.13	7.59	6.11	5.69	20.61
Depth Imbalance within $1.00 (%)	-10.53	60.87	6.32	5.47	4.20	16.54
Panel B: Large Securities (*N*=30)						
Bid-Ask Spread:						
Quoted Spread (%)	0.30	0.21	1.99	1.26	1.21	14.89
Limit Spread (%)	0.39	0.44	1.98	1.26	1.22	14.90
1,000-share Spread (%)	0.76	0.78	1.71	1.34	1.32	14.62
Limit Book Depth:						
Depth within $0.50 (000)	35.88	37.36	2.27	1.94	1.58	19.27
Depth within $1.00 (000)	54.39	53.66	3.24	1.92	1.87	23.41
Depth Imbalance within $0.50 (%)	-11.22	52.01	1.87	1.60	1.55	16.73
Depth Imbalance within $1.00 (%)	-13.76	47.55	2.16	1.81	1.65	18.75
Panel C: Small Securities (*N*=30)						
Bid-Ask Spread:						
Quoted Spread (%)	1.01	0.68	2.06	1.94	1.41	18.01
Limit Spread (%)	2.86	3.72	2.93	2.24	2.02	23.89
1,000-share Spread (%)	4.77	4.81	3.62	2.34	1.90	26.26
Limit Book Depth:						
Depth within $0.50 (000)	9.72	15.95	3.62	2.64	2.18	28.13
Depth within $1.00 (000)	14.83	23.90	4.09	3.00	2.50	32.03
Depth Imbalance within $0.50 (%)	-2.87	74.76	3.21	2.71	2.39	28.90
Depth Imbalance within $1.00 (%)	-7.75	70.01	3.80	3.10	2.25	30.91

Table 10. Explained variation in bid-ask spreads

The table lists the average R^2 from stock-specific regressions of execution costs measures on stock-specific order flow and volatility variables and order flow principal components. All variables are calculated at 15-minute intervals from 9:30 a.m. through 4:00 p.m. In addition, all variables used in the principal components analysis are standardized using firm and time-of-day specific means and standard deviations. Absolute returns are defined based on quote midpoints at the beginning and end of each period. Trades include all trade executions from TAQ. Order submissions include all market and limit order submissions through the NYSE's electronic SuperDOT system. The Limit Order Ratio is the proportion of total order submissions that are limit orders. Order flow common factors (principal components) are estimated from total market plus limit order submissions. Principal components are estimated for total order flow and for order flow categorized by trader type, where trader types include program traders, institutional traders, retail traders, and exchange members. Results are provided for three measures of execution costs. The percentage quoted bid-ask spread is the difference between the ask and bid prices divided by the quote midpoint. The percentage limit spread is the difference between the best ask and bid prices in the limit order book divided by the average of these two prices. The percentage 1,000-share spread is defined as the difference between the 1,000-share ask and the 1,000-share bid divided by the average of these two prices, where the 1,000-share ask (bid) is the price one would have to go up (down) to in order to reach 1000 shares of cumulative limit book depth. The sample includes 100 NYSE-listed securities selected randomly across market capitalization deciles and the sample period is from November 1997 through February 1998. Results for the full sample are provided in Panel A. Panels B and C provide results for subsamples categorized by market capitalization, where small and large stocks are defined as deciles 1-3 and 8-10, respectively.

Incremental Variables:	% Quoted Spread		% Limit Spread		% 1,000-share Spread	
	Cum. R^2	Incr. R^2	Cum. R^2	Incr. R^2	Cum. R^2	Incr. R^2
Panel A: All Securities (*N*=100)						
Own Absolute Return and Trades	0.0498	0.0498	0.0203	0.0203	0.0155	0.0155
Own Submissions	0.0537	0.0039	0.0253	0.0050	0.0249	0.0094
Own Submissions by Trader Type	0.0583	0.0046	0.0294	0.0041	0.0300	0.0051
Common Order Flow	0.0625	0.0042	0.0355	0.0061	0.0356	0.0056
Common Order Flow by Trader Type	0.0724	0.0099	0.0457	0.0102	0.0513	0.0157
Own Limit Ratio by Trader Type	0.0799	0.0075	0.0535	0.0078	0.0580	0.0067
Common Limit Ratio by Trader Type	0.0873	0.0074	0.0614	0.0079	0.0666	0.0086
Panel B: Large Securities (*N*=30)						
Own Absolute Return and Trades	0.0513	0.0513	0.0315	0.0315	0.0282	0.0282
Own Submissions	0.0523	0.0010	0.0334	0.0019	0.0336	0.0054
Own Submissions by Trader Type	0.0557	0.0034	0.0379	0.0045	0.0395	0.0059
Common Order Flow	0.0587	0.0030	0.0410	0.0031	0.0419	0.0025
Common Order Flow by Trader Type	0.0649	0.0062	0.0484	0.0074	0.0480	0.0061
Own Limit Ratio by Trader Type	0.0697	0.0048	0.0551	0.0067	0.0556	0.0076
Common Limit Ratio by Trader Type	0.0738	0.0041	0.0602	0.0051	0.0596	0.0040
Panel C: Small Securities (*N*=30)						
Own Absolute Return and Trades	0.0351	0.0351	0.0117	0.0117	0.0078	0.0078
Own Submissions	0.0403	0.0052	0.0179	0.0062	0.0177	0.0099
Own Submissions by Trader Type	0.0454	0.0051	0.0223	0.0044	0.0239	0.0062
Common Order Flow	0.0514	0.0060	0.0335	0.0112	0.0342	0.0103
Common Order Flow by Trader Type	0.0659	0.0145	0.0470	0.0135	0.0564	0.0222
Own Limit Ratio by Trader Type	0.0752	0.0093	0.0548	0.0078	0.0634	0.0070
Common Limit Ratio by Trader Type	0.0857	0.0105	0.0673	0.0125	0.0780	0.0146