Article can be accessed online at http://www.publishingindia.com

Understanding Order-Flow Volatility

Rahul Ravi

Abstract

Given that order-flow is likely to be driven by differences in investors' beliefs, a reasonable hypothesis is that order-flow volatility should be positively related to the level of investor heterogeneity. Motivated by this hypothesis, this study investigates the association between order-flow variability and various known proxies of divergence of opinions and informational differences. We find order-flow variability to be positively associated with trading volume, dispersion in analysts' forecasts and the S&P 500 futures open interest (a proxy for market-wide divergence of opinions), and negatively associated with the adverse selection cost of trading. We also demonstrate a positive relation between order-flow variability and risk-adjusted stock returns. In conclusion, we find evidence of co-movement in order-flow variability as well as in the adverse selection cost of trading and liquidity. Co-movement in order-flow variability appears to partially explain co-movement in liquidity.

Keywords: CAPM, Beta, Markowitz Mean-Variance Framework, Fama and French

Introduction

Order-flow refers to the arrival of buy and sell orders in the market. Volatilityof order-flow therefore suggests varying levels of buying and selling pressures in the market. To the extent that this variability could affect asset returns (and possibly also the functioning of the financial market itself), at least in the short term; it is important to understand its drivers. This study attempts to understand volatility of order-flow by exploring its link to factors such as information asymmetry and divergence of opinions among investors in the market.

Microstructure theory suggests various reasons why order-flow volatility should be related to asset returns. Kyle (1985) shows that the variability of uninformed orders affects the extent to which order-flow moves prices – the price impact parameter. Similarly, inventory models such as the one devised by Ho and Stoll (1981) suggest that a market maker's inventory costs will increase in the variability in buy and sell orders. Both arguments suggest that the volatility of order flow will affect the level of trading costs (albeit, in different directions) and thereby should be related to expected returns (Amihud and Mendelson, 1986; Brennan and Subrahmanyum, 1996). Similarly, the literature focusing on differences of opinion (Miller, 1977) also suggests that greater heterogeneity among investor population (and thus greater order-flow variability) will, in the presence of short selling constraints, lead to inflated prices and to subsequent corrections, thus affecting returns (at least in the short term).

Motivated by these arguments, this study investigates the relation between the volatility of order flow (SIGOF) and stock returns, the bid-ask spread and its components. We use transaction level data to construct a fifteen-minute time series of net order-flow (Total buy orders minus total sell orders) of each stock for a sample of 5418 NYSE listed stocks, spanning across thirteen years (January 1993 through December 2005). We estimate monthly SIGOF for a stock as the standard deviation of this series within each month. On average, the sample consists of 1,730 firms in every calendar month.

We find that, on an average, higher SIGOF is associated with lower per dollar adverse selection cost of trading, lower inventory cost (per dollar), higher trading volume, and higher dispersion in analysts' forecasts. The negative relationship between order-flow variability and inventory costs is somewhat puzzling. The relationship between

* Assistant Professor, Department of finance, John Molson School of Business, Concordia University, Canada.
Email: rravi@jmsb.concordia.ca

adverse selection cost and order-flow variability may be understood using the Kyle (1985) framework, whereby, more variable order-flow allows informed investors to hide their trade more effectively. The positive relation between SIGOF and trading volume is consistent with the predictions of the divergence of opinion literature. We also find that increase in SIGOF is associated with an increase in risk adjusted return, increase in trading volume and an increase in the number of S&P open interest contracts. These results are consistent with the assertions ofthe divergence of opinion literature.[1]

In conclusion, this study attempts to throw some light on the issue of possible commonality in order-flow volatility. Several papers study common effects in order imbalance (Hasbrouck and Seppi, 2001; Harford and Kaul, 2005). Our interest in common effects and order-flow variability is motivated by the idea that differences of opinions could be systematic rather than stock specific (Miller, 1977). We present evidence of significant commonality in order-flow variability, with 83% of the stocks in our sample tending to move in the same direction. Building on the work of Chordia*et al.*(2001), we show that the adverse selection and inventory components of the spread display commonality. We link this commonality to co-movement in SIGOF and provide some evidence that the commonality in liquidity and transaction costs (adverse selection and inventory) is at least partially determined by the same factors that determine commonality in order-flow variability. This study further provides some suggestions as to the identity of these factors by demonstrating that SIGOF contains a systematic component, plausibly associated with aggregate divergence in opinions.

The remainder of this paper is organized as follows. The next section develops our hypothesis. The third section describes the construction of the key variables. The fourth section details the data and the sample. The next section presents the results and the last section offers some conclusions.

Hypotheses Development

To the extent that the lack of agreement among investors with respect to the value of a stock is likely to manifest in the market in the form of order-flow variability, we argue that SIGOF presents a plausible measure of investor heterogeneity. This section develops several hypotheses to test the above assertion.

SIGOF and Trading Costs

Models such as Kyle (1985) demonstrate that trading costs increase in the degree of the potential information asymmetry between the market maker and the informed investors. Nevertheless, *ceteris paribus,* trading costs should decline in the variability of uninformed trading as they allow the informed trader to hide trades more effectively.

H_1: The adverse selection cost of trading should be negatively related to the variability of order-flow.

Other models such as Ho and Stoll (1981), which focus on inventory costs of the market maker, suggest that asynchronous timing in buy and sell orders imposes inventory management cost on the market-maker. Therefore, a market maker's inventory costs should increase in the variability of order imbalance.

H_2: The inventory holding cost of the market maker should be positively associated with the variability of order-flow.

SIGOF and Trading Volume

Why do investors trade such enormous quantities? Differences in information alone cannot explain high levels of trading volume (Milgrom and Stokey, 1982). Harris and Raviv (1993) and Kandel and Pearson (1995) show that differences in opinions help to explain the high levels of trading volume, and that a greater divergence in opinion leads to higher trading volume. These differences can arise either due to differences in prior beliefs or due to differences in the way investors interpret public information.

H_3: Trading volume should be positively correlated with contemporaneous order-flow variability.

Anshuman, Chordia, and Subrahmaniam (2001) explore the properties of variability in trading volume (s_{Vol}), and suggest that s_{Vol} can be interpreted as a measure of divergence in opinions.

[1] We also find positive association between SIGOF and some commonly used proxies for divergence in opinions, namely: market capitalization, S&P 500 futures open interest, dispersion in analyst forecasts, and the volatility of trading volume.

H_4: Variability in trading volume should be positively correlated with contemporaneous order-flow variability.

SIGOF and Dispersion in Analyst Forecasts

Diether, Malloy, and Scherbina (2002) use dispersion in analyst annual earnings forecasts as a proxy for differences in opinions. They find that stocks with higher dispersion in analysts' earnings forecasts earn significantly lower future returns than do otherwise similar stocks. If order-flow variability is a measure of differences in opinions, we should see a positive contemporaneous relation between SIGOF and dispersion in analyst forecasts.

H_5: Dispersion in analyst forecasts should be positively correlated with contemporaneous order-flow variability.

SIGOF and Returns

Miller (1977) suggests that short-sale constraints prevent pessimistic opinions from being fully reflected in stock prices. Miller argues that a stock's price will reflect the valuations of optimistic investors because pessimists cannot participate in the market when short sale constraints are in place.[2] Thus, in the presence of short sale constraints, stocks may become overpriced during periods of high differences of opinions about their prospects. Therefore, to the extent that SIGOF is related to the level of divergence in opinions, we should expect to find a positive contemporaneous relation between SIGOF and stock returns.

H_6: Order-flow variability should be positively associated with contemporaneous returns.

SIGOF and Market-Wide Divergence in Opinions

Miller (1977) argues that the divergence in opinions is not entirely idiosyncratic, but is correlated with both the systematic and the non-systematic components of a stock's return. While the above set of hypotheses relates to idiosyncratic differences in opinions, the next hypothesis explores the relationship between SIGOF and systematic divergence in opinions. Bessembinder, Chan, and Seguin (1996) propose a useful proxy for systematic dispersion in opinions, suggesting that open interest in the S&P 500 index futures contract captures the cross-sectional dispersion in traders' opinions regarding the market-wide prospects. We accordingly relate the time-series of SIGOF of each stock in the sample to open interest in the S&P 500 index futures contract.

H_7: On average, SIGOF should be positively related to the open interest in the S&P 500 index futures contract.

Co-Movement in SIGOF and Liquidity

The final section of this study attempts to explore the levels of co-movement in order-flow variability. Drawing upon the arguments leading to Hypothesis 7, the systematic component in SIGOF should induce co-movement in order-flow variability across stocks. We take this argument further by suggesting that if the spread and the components of the spread are related to order-flow variability (H1 and H2), then co-movement in SIGOF is likely to induce co-movement in the spread and in its adverse selection and/or in the inventory components.

H_8: Co-movement in order-flow variability will induce co-movement in the spread, the adverse selection component and the inventory component of the spread.

Construction of Variables and Empirical Methods

We carry out the empirical analysis in three stages. We start with a firm-level contemporaneous analysis to study the association between SIGOF and the adverse selection cost per dollar traded (DVIA), the inventory cost per dollar traded (DVINV), risk-adjusted stock returns, trading volume (vol), market capitalization (Size), variability of trading volume (s_{Vol}), dispersion in analysts' forecasts (DISP), and the number of analysts following a stock (ANAL). In the second stage, we explore the determinants and the effects of SIGOF. Finally, we examine co-movement in SIGOF and its implications for co-movement in liquidity.

[2] This argument is reinforced by the fact that arbitrage is risky and costly (e.g., Pontiff, 1996).

Measuring Order-Flow Variability

We divide each trading day (9:30 a.m. to 4:00 p.m.) in a given month into 26 15-minute intervals. For every stock in our sample, we compute order-flow (the number of buyer-initiated trades minus the number of seller-initiated trades) in each interval.[3] For a given stock x, $SIGOF_{x,t}$ is the standard deviation of the 15-minute order-flow series in month t. We compute two additional measures of order-flow variability based on alternative definitions of order-flow: first, as the difference between the volume of buyer-initiated trades and the volume of seller-initiated trades; and second, as the difference between the value of buyer-initiated trades and the value of seller-initiated trades. The results are identical, and therefore, for the sake of brevity, this paper only discusses the results corresponding to the number of trades-based measures of order-flow. Moreover, while the volume-based measure is likely to be contaminated by volume effects, the value measure is likely to be affected by prices. Since several of the hypothesized variables are functions of either price or volume, we believe the choice of number of trades-based measure of order-flow is more conservative.

Measures of Liquidity

We use the quoted spread and proportional quoted spread as two related measures of liquidity. The quoted spread (QSPR) is defined as $QSPR = (P_A - P_B)$ where P_A is the ask price and P_B is the bid price. Defining the quote midpoint as $P_M = (P_A + P_B)/2$, the proportional quoted spread is defined as $PQSPR = (P_A - P_B)/P_M$.

The Adverse Selection and Inventory Cost Components of The Spread

We estimate the components of the bid-ask spread using the method advocated by Lin, Sanger, and Booth (LSB, 1995).[4] This method is based on the approach described in Stoll (1989) and related to the approach used by Huang and Stoll (1997). LSB use a regression approach to estimate the proportion of the effective spread that can be attributed to information asymmetry. The basic idea is that the quote revision reflects the adverse selection component of the spread, while the change in the transaction price reflects the order processing costs and bid-ask bounce.

In the LSB model, information revealed by the trade at time t is reflected in the quote revisions. If P_t is the transaction price at time t, and Q_t is the quote midpoint at time t, then $B_t = B_{t-1} + \lambda S_{t-1}$ and $A_t = A_{t-1} + \lambda S_{t-1}$, where B_{t-1} and A_{t-1} are the prevailing bid and the ask prices at time t. λ can be interpreted as the proportion of the effective spread due to adverse selection. $S_{t-1} = P_{t-1} - Q_{t-1}$ is one-half of the effective spread. The revision in the quote mid-point is expressed as

$$\Delta Q_t = \lambda S_{t-1} + \varepsilon_t \tag{1}$$

$$S_t = \theta S_{t-1} + \eta_t \tag{2}$$

where $\Delta Q_t = Q_t - Q_{t-1}$ and $Q_t = \frac{(B_t + A_t)}{2}$. θ represent the order processing cost component of the spread, and $(1 - \lambda - \theta)$ represents the inventory component of the bid-ask spread. We calculate the per dollar adverse selection cost of trading (DVIA) by multiplying λ by the average monthly effective spread and dividing it by the average transaction price for the month. We use the same method to calculate the per dollar inventory cost of trading (DVINV).

Other Variables

Trading volume (VOL) is the total number of shares traded in each month, as reported in the Center for Research in Security Prices (CRSP) database. The standard deviation of trading volume (s_{Vol}) is estimated in a manner analogous to the SIGOF measure. We define s_{Vol} as the standard deviation of 15-minute trading volume, calculated across all 15-minute intervals in a given month.

Monthly holding period returns (r) are obtained directly from the CRSP monthly tapes. We calculate the risk-adjusted return using the four-factor model of Carhart (1997).This model is an extension of the Fama and French (1993) three-factor model, incorporating an additional momentum factor. For each month, we run the following regression for firms with more than 17 daily return observations within that month,

[3] Trades are classified into buyer or seller initiated trades using the procedure laid out in Lee and Ready (1991).

[4] We have also run our analysis using the adverse selection components proposed by Glosten and Harris (1988) and Neal and Wheatley (1998) as well as the price impact parameter based on the Hasbrouck (1991). Our results remain qualitatively unchanged and are found to be robust to the method selected. For the sake of brevity, we only report the results corresponding to LSB (1995).

$$R_{i,t,d} = \alpha_{i,t} + \beta_{i,m} \times r_{m,d} + \beta_{i,SMB} \times SMB_{t,d} + \beta_{i,HML} \times HML_{t,d} + \beta_{i,MOM} \times MOM_{t,d} + \varepsilon_{i,t,d} \tag{3}$$

where, for day *d* in month *t*, $R_{i,t,d}$ is stock i's excess return; $r_{m,d}$ is the excess return on the market portfolio; and $SMB_{t,d}$ and $HML_{t,d}$ are the Fama-French (1993) size and book-to-market portfolios. $Mom_{t,d}$ is the momentum factor in month*t*. $\varepsilon_{i,t,d}$ is the residual with respect to the described factor model. The data for the three factors (*HML, SMB, Mom*) are obtained from Ken French's website.[5] The risk-adjusted return for stock *p* in month *t* is calculated as ($\alpha_{i,t}$). The daily risk-adjusted return is given by $r_{i,t,d} = \alpha_{i,t} + \varepsilon_{i,t,d}$.

Higher market-to-book ratio has often been interpreted as indicative of higher growth options in the firm. To the extent that growth options are relatively more difficult to value, firms with higher market-to-book ratio are likely to have higher investors' heterogeneity. The market-to-book-ratio of the firm (*MB*) is calculated as:

$$MB = \frac{(Common\ shares\ outstanding) \times (Share\ Price) + (Total\ assets) - (Common\ equity)}{(Total\ assets)}$$

Since larger firms are likely to attract a broader cross-section of investors, they are also more likely to be affected by greater investors' heterogeneity. Size is defined as the month-end shares outstanding, times the month-end closing price. Dispersion in analysts' forecasts (DISP) is used as a proxy variable for divergence in opinion and is measured as the standard deviation of current fiscal year earnings forecasts, divided by the consensus mean of current fiscal year earnings forecasts. The DISP data are obtained from I/B/E/S.We use the number of analysts providing forecasts (ANAL) as a control variable.

The variables for inclusion in exploring the characteristics of a stock's order-flow variability are discussed in the second section. Hypotheses 1 through 5 refer to cross-sectional associations involving order-flow variability. To test these hypotheses, we use the following cross-sectional regression model:

$$SIGOF_{i,t} = \alpha_t + \beta_{1,t} \times \ln(\sigma_{VOL,i,t}) + \beta_{2,t} \times \ln(Vol_{i,t}) + \beta_{3,t} \times \ln(Size_{i,t}) + \beta_{4,t} \times IA_{i,t} + {} + \beta_{5,t} \times INV_{i,t} + \beta_{6,t} \times \ln(Anal_{i,t}) + \beta_{7,t} \times DISP_{i,t} + \beta_{8,t} \times MB_{i,t} + \varepsilon_{i,t} \tag{4}$$

where the subscripts (i,t) denote stock *i*and month *t*. Size corresponds to the market capitalization of the firm; *Vol* is monthly trading volume; and s_{Vol} is the variability in trading volume. *ANAL* is the number of analysts following the firm. *DISP* is the dispersion in analyst forecasts. *MB* is the market-to-book ratio of the firm. The number of analysts' providing forecasts (ANAL)are added to the model as a control variables.

Hypotheses 6 and 7 refer to time-series associations involving order-flow variability. To test these, we require the use of a time-series regression model. Several of the variables used in this analysis are persistent through time, which raises concerns about the inference of causality and contemporaneous associations. We get around this problem by following Chordia, Roll, and Subrahmanyam (2001) in using monthly proportional changes in the variables, rather than levels in the regression analysis. For example, for the variable M, the proportional change is defined as $(M_t - M_{t-1})/M_{t-1}$. The model used is as follows:

$$DSIGOF_{i,t} = \alpha_i + \beta_{1,i} \times DNOIC_t + \beta_{2,i} \times DSize_{i,t} + \beta_{3,i} \times DVol_{i,t} + \beta_{4,i} \times DSigVol_{i,t} + \beta_{5,i} \times r_{i,t} + \beta_{6,i} \times DANAL_{i,t} + \beta_{7,i} \times DDISP_{i,t} + \beta_{8,i} \times DMB_{i,t} + \varepsilon_{i,t} \tag{5}$$

The subscripts (i,t) refer to stock *i* and month *t*, respectively. *D* denotes proportional change and the subscript *t* indicates that the change is being calculated between trading months*t-1* and *t*. *NOIC* is S&P 500 futures open interest (measured as number of contracts), *DSigVol* is the proportional change in s_{Vol}.

Co-Movement in SIGOF

This section develops two related methods to explore the existence of commonality in order-flow volatility and thus test the assertions of Hypothesis 8. First, we use pair-wise correlation analysis. We estimate the pair-wise correlation between the quoted spread for each of the 5,418 firms in the sample (Corr ($QSPR_{i,t}$, $QSPR_{j,t}$)) for all $i \neq j$. To assess the role of *SIGOF*, we orthogonalize the quoted spread for firm *i* with respect to its SIGOF:

$$QSPR_{i,t} = \alpha_i + \beta_i \times SIGOF_{i,t} + \varepsilon_{i,t} \tag{6}$$

The residuals $\varepsilon_{i,t}$ represent the spread for firm *i* in month *t*, while controlling for the order-flow volatility of firm *i*. We re-estimate the pair-wise residual correlation $Corr(\varepsilon_{i,t}, \varepsilon_{j,t})$. Comparing $Corr(QSPR_{i,t}, QSPR_{j,t})$ with $Corr(\varepsilon_{i,t}, \varepsilon_{j,t})$ provides a way of quantifying the contribution of *SIGOF* to spread co-movement. We repeat the analysis using proportional spreads as a related measure of liquidity.

The second method for exploring co-movement in liquidity

[5] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

and the role of *SIGOF* in such co-movement is closely related to the work of Chordia, Roll, and Subrahmanyam (2001). We estimate time-series regressions relating monthly proportional changes in *SIGOF* for individual stocks to market-wide average order-flow variability, i.e.,

$$DSIGOF_{i,t} = \alpha_i + \beta_i \times DSIGOF_{M,t} + \gamma_{i,1} r_{m,t-1} + \gamma_{i,2} r_{m,t} + \gamma_{i,3} r_{m,t+1} + \gamma_{i,4} \ln\left(\frac{P_{i,t,\max}}{P_{i,t,\min}}\right) + \varepsilon_{i,t}$$

where $DSIGOF_{i,t}$ is the proportional change in order-flow variability for stock i from trading month t-1 to t. $DSIGOF_{M,t}$ is the corresponding change in market-wide *SIGOF*, calculated as:

$$SIGOF_{M,t} = \frac{\sum_{i=1}^{n} SIGOF_{i,t}}{n} \quad \text{....................... (7a)}$$

$$DSIGOF_{M,t} = \frac{\left(SIGOF_{M,t} - SIGOF_{M,t-1}\right)}{SIGOF_{M,t-1}} \quad \text{................ (7b)}$$

$r_{m,t+1}$, $r_{m,t}$, and $r_{m,t-1}$ are the lead, contemporaneous, and lag market returns, respectively, included in the model to control for any possible effects of returns on order-flow variability.[6] The contemporaneous natural logarithm of the ratio of the maximum and minimum prices of stock i in month t is included as a control for volatility. The β_i coefficients may be interpreted as a measure of co-movement in *SIGOF*.

In computing the market index $DSIGOF_{M,t}$, the value of stock *i* is excluded, and thus, the explanatory variable in the above regression is slightly different for each stock's time-series regression. We estimate model (8) for the two measures of liquidity (*QSPR*and *PQSPR*), the adverse selection cost per dollar traded (*DVIA*), and the inventory cost per dollar traded (*DVINV*).

To explore the role of *SIGOF* in liquidity co-movement, we control for the effect of *SIGOF* on *QSPR*, *PQSPR*, *DVIA*, and *DVINV*, using the OLS specification:

$$M_{i,t} = a_i + b_i SIGOF_{i,t} + \varepsilon_{i,t} \qquad (8)$$

We then repeat the co-movement analysis in Equation (7) on $\varepsilon_{i,t}$.

6 Estimated the average market SIGOF (equations 7a, and 7b), is adjusted by removing firm i effect from it before using it to estimate model 7.

Sample Selection and Sample Characteristics

The sample period runs from January 1993 to December 2005.[7] Data were retrieved from the NYSE Trade and Quote (TAQ), Compustat, and Center for Research in Security Prices (CRSP) databases. Analyst data were obtained from the I/B/E/S database. Utilities (SIC code 49 to 50), and firms from the financial sector (SIC code 60 to 68), were excluded because these are regulated industries. ADRs, other securities incorporated outside the US, as well as preferred stocks and other non-common stocks, were excluded.[8] We delete all non-NYSE firms from the sample.[9]

Several filters were employed to ensure the validity of the TAQ data.[10] The first trade of each day is dropped from the analysis, since it usually occurs through a call auction. The TAQ database does not eliminate auto-quotes (passive quotes by secondary market dealers), which may cause the quoted spreads to be artificially inflated. Since no reliable method can exclude auto-quotes in TAQ, only BBO (best bid or offer) eligible primary market (NYSE) quotes were used (Chordia, Roll, and Subrahmanyam

7 We decided to limit our sample to 1993 till 2005 because TAQ starts in 1993 and using data beyond 2005 exponentially increases the volume of data while not necessarily adding anything towards the objective of this study.

8 Securities with CRSP share codes different from 10 or 11 were excluded.

9 The spread decomposition methodologies used in this paper are appropriate for a specialist market (NYSE), as opposed to dealer markets (NASDAQ). In addition, interpretation of the spread components for NASDAQ trade and quotes is problematic due to the presence of inter-dealer trades in the data. These non-informational trades cannot be identified in the database. Restricting this study to NYSE-based firms also abstracts from differences in market structure.

10 We drop all trades with a correction indicator other than 0 or 1, and retain only those trades for which the condition is B, J, K, or S.We also drop all trades with non-positive trade size or price.Finally, we omit all trades recorded before opening time or after the closing time of the market. Negative bid-ask spreads and transaction prices are also eliminated. In addition, only quotes that satisfy the following filter conditions are retained: we eliminate all quotes for which the quoted spread is greater than 20% of the quote midpoint, when the quote midpoint is greater than $10 or when the quoted spread is greater than $2, when the quote midpoint is less than $10. We also eliminate all quotes for which either the ask or the bid moves by more than 50%.

2001, 2002).[11] The trade and the quote data are matched following Lee and Ready (1991).

The adverse selection (*DVIA*), and the inventory (*DVINV*) components of the spread, as well as the order-flow variability (SUGOF), are generated from the TAQ data. The sample size ranges from a minimum of 1,605 firms in January 1993 to a maximum of 2,194 firms in April 1998, and to 2,089 in December 2005. The full sample consists of a total of 1,354,900,396 matched trade and quote pairs. Using these pairs, we compute monthly *DVIA, DVINV*, and *SIGOF* for each firm in the sample period. Monthly volume, size, and return data from the CRSP; market-to-book ratio is calculated using quarterly Compustat data. The number of analysts and the dispersion in analyst earnings forecasts is from I/B/E/S. Our final dataset consists of 188,304 firm-months of data.

Results

Table 1 presents the distribution of firms over the sample period, and some descriptive information. The size of the average firm in the sample increases from 1993 to 2000, and then drops for the rest of the study period. The average market-to-book ratio remains stable and the average number of analysts following a firm declines over the sample period. The average adverse selection cost of trading, expressed as a percentage of the quoted spread, shows a marginally increasing trend from 1993 to 1998 and then stabilizes for the rest of the sample period. Average trading volume increases almost four-fold between 1993 and 2005. The mean inventory cost component of the spread shows an almost monotonic decline, with a minimum of 9.42% in 2005.

Exploring Order-Flow Variability

Table 2 presents the time-series distribution of the mean *SIGOF* across industries.[12] While the exact ranks of industries with respect to average *SIGOF* changes from year to year, high tech industries such as health care, drugs and genetic engineering, and computer manufacturing have relatively more volatile *SIGOF*, while industries such as wholesale and construction display relatively lower order-flow. The former group consists of high growth firms with significantly large proportion of intangible assets. These firms would be relatively difficult to value vis-à-vis the firms in the latter group. The patterns in Table 2 provide some preliminary support for our interpretation of *SIGOF* as a measure of the level of investors' heterogeneity or the extent of less than perfectly informed trading in the market.

Table 3 shows the time-series and cross-sectional variation in average *SIGOF*, sorting at the end of each month by firm size (Panel A), market-to-book ratio (Panel B), and the dispersion in analysts' earnings forecasts (*DISP*) (Panel C). Average *SIGOF* increases as each of the three variables increase. The trend is consistent across years. These results support the *heterogeneity* interpretation of SIGOF.

Average *SIGOF* is fairly stable until 1996, and increases monotonically from 1996 to 2005 (Table 1). Table 3 suggests that this time-series pattern is most prominent for the largest 20% of the firms and relatively weak among the smallest 20% of the firms. The same time-series pattern in *SIGOF* can also be seen with respect to the dispersion in analysts' earnings forecasts (*DISP*). The time-series pattern in *SIGOF* is also present across market-to-book quintiles, though in this case, it is more prominent among the higher quintiles.

Table 3 suggest that larger firms, as well as firms with more growth options, and firms with more dispersed earnings forecasts, tend to have high order-flow variability. To the extent that these characteristics are not orthogonal to each other, these results need to be interpreted with caution. Table 4 attempts to further explore the results in Table 3 by examining *SIGOF* for two-way sort. We divide the sample into *size* and *MB* quintiles (Panel A), *MB* and *DISP* quintiles (Panel B), and *size* and *DISP* quintiles (Panel C).

Controlling for size, the book-to-market effect observed in Table 3 becomes much weaker. Order-flow variability is low for smaller firms and high for larger firms. Similarly, when controlling for the dispersion in analysts' earnings forecasts, the market-to-book effect becomes, once again, considerably weaker. A possible explanation for the weakening of the market-to-book effect could be related to the ambiguous nature of the variable. While high market-to-book is usually interpreted as indicating high growth opportunities, it could also signal overvalued firms. Panel C of Table 4 stratifies the sample by *DISP*

[11] All quotes with condition 5, 7, 8, 9, 11, 13, 14, 15, 16, 17, 19, 20, 27, 28, 29 were excluded.

[12] We use an adapted version of the 14-industry classification, as proposed in Ritter (1991).

and Size. *SIGOF* is low for small firms with low earnings uncertainty across analysts following them and high for large firms with relatively more uncertain earnings.

The results, so far, consistently suggest a positive association between *SIGOF* andthe level of heterogeneity among investors. The next set of results explores this association in more detail. Table 5 summarizes the Spearman correlation matrix of the key variables (Pearson correlations give similar results). The table provides the average monthly correlation coefficients, obtained by estimating cross-sectional correlations for every month from January 1993 to December 2005 and calculating their time-series averages.

The average pair-wise correlation between *SIGOF* and the per dollar adverse selection cost is -0.572. This lends some preliminary support to Hypothesis 1. The negative correlation between *DVIA* and *SIGOF* suggests that, in times of high order-flow variability, the market maker is less concerned about losses due to adverse selection. The average pair-wise correlation between *SIGOF* and trading volume (*VOL*) is 0.660. A positive correlation between *trading volume* and *SIGOF* suggests that, on average, firms with more volatile order-flow will experience greater trading volume. The positive correlation between *trading volume* and *SIGOF* provides evidence in support of Hypothesis 3. The correlation between variability in trading volume (s_{Vol}) and order-flow variability (*SIGOF*) is 0.488. This lends preliminary support to Hypothesis 4.

Table 5 also provides some evidence in support of Hypotheses 5 and 6. The correlation between dispersion in analysts' forecasts (*DISP*) and *SIGOF* is 0.036, which, though small, is statistically significant. This result suggests that firms with more volatile order-flow are also more likely to have more dispersed analysts' earnings forecasts. The positive and significant correlation between *SIGOF* and risk-adjusted returns (*r*) lends support to Hypothesis 6. This implies that periods of high order-flow variability are likely to be associated with higher returns.

We estimate the model specified in Equation (4) by event month.Table 6 presents the time-series averages of the estimated coefficients and the t-statistics corresponding to the test $H_0: \bar{\beta}_j = 0$.[13] We find a negative and significant association between *adverse selection cost of trading* and *SIGOF*. This reinforces the support for Hypothesis 1. We find evidence of negative association between *inventory management cost* and *SIGOF*. This result contradicts Hypothesis 2. A possible explanation for this seemingly puzzling relation could be the limitations of the three-way decomposition model used in estimating the inventory cost variable.[14] The positive and significant coefficients on *Vol* and s_{Vol} lend support to Hypotheses 3 and 4. The results suggest that periods of high volume and volume volatility are associated with high SIGOF. The positive association between dispersion in analyst forecasts and SIGOF is consistent with hypothesis 5.

The coefficient for the number of analysts following the firm is found to be consistently positive. This suggests that firms with more analysts following are also firms with more volatile order-flow. This result can be interpreted in at least two ways. First, a larger number of analysts generates more firm-related data and, therefore, makes more information available for smart investors to trade on making the investor pool relatively more heterogeneous and the order-flow more volatile. Second, interpretation could be that larger numbers of analysts are attracted to the stocks with high divergence of opinions because more demand exists for information in these markets. In conclusion, the results in Table 6 provide two key insights into understanding order-flow variability. First, informed traders are likely to be able to hide their trades more effectively during periods of high order-flow variability (Stealth trading). Second periods of high order-flow variability seems to be associated with periods of high divergence of opinions among investors.

Table 7 presents the results of the time-series model specified in Equation (5). We estimate the model for every stock in the sample. This table reports the cross-sectional averages of the estimated coefficients and the t-statistics corresponding to the test $H_0: \bar{\beta}_j = 0$. We find a positive association between the proportional changes in risk-adjusted returns and the proportional changes in *SIGOF*. This result is in concurrence with the predictions of hypothesis 6, based on the divergence in opinions interpretation of *SIGOF*. As the divergence in opinions increases, it drives out a fraction of the pessimists from the market, thereby inflating the stock price, and leading to higher returns.

[13] Throughout this paper, all standard errors and associated t-statistics are calculated using a Newey-West correction with four lags.

[14] While this limitation of the model would also make the adverse selection results suspect, we do find consistent results using several related and unrelated measures of adverse selection (details in footnote 4).

We also find evidence of a positive association between changes in S&P 500 futures open interest and the changes in the order-flow variability for the average stock (Hypothesis 7). Bessembinder, Chan, and Seguin (1996) suggest that open interest on the S&P 500 futures contract represents an empirical proxy for cross-sectional dispersion in traders' opinions about market information. Although statistically non-significant, the positive coefficient in Table 7 hints that the average stock's *SIGOF* could potentially contains a market-wide component, and points to the possible existence of commonality in *SIGOF*.

Commonality in Order-Flow Variability

The arguments developed in section three suggest existence of systematic factors driving order-flow volatility, which in turn should lead to commonality in SIGOF. Although the results in table 7 fail to find any statistical support for this assertion, it may possibly be attributed to noisy or inaccurate proxies. This section uses the methodology described in Section 'Co-movement in SIGOF'to explore for this commonality and its implications for co-movement in liquidity, adverse selection costs, and inventory costs.[15]

Following the methodology outlined in Section 'Co-movement in SIGO', Table 8 presents the statistics for the β_i coefficients from Equation (7). Over 87% of the individual β_i are positive, with over 49% significant at the 5% one-tailed critical value. For the quoted spread, β_i is positive for approximately 93% of the stocks in the sample; 67% of these are statistically significant. The corresponding proportion of positive (positive and significant) β_i coefficients for proportional spreads, adverse selection costs, and inventory costs are 96% (78%), 93% (58%) and 77% (27%), respectively. These results provide evidence of co-movement in *SIGOF*, *liquidity*, *adverse selection costs,* and *inventory costs* (consistent with Chordia, Roll, and Subrahmanyam, 2001). The average R^2 for the regressions are about 12% for the quoted spread, 16% for the proportional quoted spread, 6% for *SIGOF*, 6% for *DVIA*, and about 2% for *DINV*.

Table 9 (Panel A) presents the results of estimating Equation (7), using the level of each variable instead of the proportional change. For *SIGOF*, 86% of the β_i coefficients are positive and 75% are positive and significant. In the case of the proportional spreads, we find that 92% of β_i are positive while 84.5% are positive and significant. Of the coefficients for the monthly quoted spreads, 95% are positive while 91% are positive and significant. Of the coefficients for the adverse selection cost component, 92% are positive, and 88% of the inventory cost component coefficients are also positive. These results are similar to the findings noted in Table 8 (based on changes). Table 9 provides additional evidence in support of the existence of systematic components of *SIGOF* contributing to co-movement in *SIGOF*.

Next, this study looks at the implications of co-movement in *SIGOF* for co-movement in *liquidity, adverse selection costs,* and *inventory costs*. Table 9 (Panel B) provides the results of estimating Equation (7), using the residuals from Equation (8). Comparing the statistics in Panel A with those in Panel B allows us to identify the contribution of *SIGOF* in *QSPR*, *PQSPR*, *DVIA,* and *DVINV* co-movement. The explanatory power of the regressions in Panel B is lower than those in Panel A. The adjusted R^2 declines from 63% to 27% for quoted spreads (*QSPR*), from 49% to 24% for proportional spreads (*PQSPR*), from 28% to 14% for *DVIA*, and from 34% to 16% for *DVINV*. These results suggest that commonality in liquidity and in trading costs are at least partially determined by order-flow variability, or factors determining order-flow variability. These results are in favour of Hypothesis 8.

The results of the pair-wise correlation analysis, exploring the contribution of *SIGOF* co-movement to co-movement in liquidity, are presented in Table 10. The average (median) pair-wise correlation between quoted spreads is 0.5413 (0.6029). Controlling for the contemporaneous *SIGOF* (Equation 6), we find that the magnitude of correlation drops to a mean (median) level of 0.2200 (0.2720). We find a similar decline in the pair-wise correlation for proportional spreads, where the mean (median) correlation drops from 0.4620 (0.4373) to 0.1190 (0.2815). The cross stock mean (median) correlation in the adverse selection cost declines from 0.2064 (0.2250) to 0.0586 (0.0374), and the correlation in inventory cost drops from 0.2594 (0.2173) to 0.1721 (0.0695). These results provide evidence in support of Hypothesis 8, whereby order-flow volatility (or factors driving it) can partially explain co-movement in liquidity, adverse selection costs, and inventory costs.

[15] Chordia, Roll, and Subrahmanyam (2001) present arguments supporting co-movement in inventory and adverse-selection costs.

Table 11 attempts to explore the role of divergence of opinions in generating the commonality in order-flow variability. We run co-movement analysis (as described in Equations 7 and 8) on *SIGOF*. The following set of lagged variables are used as equation 8 controls: *SIGOF*, volume volatility (s_{Vol}), S&P 500 open interest (*NOIC*), risk-adjusted market returns, number of analysts (*ANAL*), market-to-book ratios, and dispersion in analysts' forecasts (*DISP*). The first column in Table 11 presents the co-movement in *SIGOF* (base case). The remaining columns correspond to co-movement analysis using Equation 8 residuals, controlling for the various proxies ofinvestors' heterogeneity. As we control for the effects of various proxies, we notice a decline in β_i from 0.834 to 0.244. The t-statistics declines from 35.5 for the base case to 7.403. The adjusted R^2 declines from 46.60% to 0.60%. The percentage of stocks with positive β_i also declines from 85.67% to 70.95%. These results suggest that divergence in opinions is, at least partially, responsible for the observed commonality in the order-flow variability, and hence, possibly for the commonality in liquidity as well as for the adverse selection and the inventory costs of trading.

Conclusion

This paper explores order-flow variability and examines the relationship between order-flow variability and divergence of opinions among heterogeneous investor pool. Our results suggest that order-flow variability is positively related to divergence of opinions among investors and negatively related to the level of information asymmetry across investors. This paper refersto 'heterogeneity among investors' in a rather broad sense, in referring to the dispersed beliefs of traders in the market. This dispersion could result from either divergence in opinions or differences in information endowment across investors.We also find that order-flow variability is not purely idiosyncratic. In other words, it exhibits commonality across stocks. The second part of this study attempts to explore this commonality in order-flow variability and link it to previously shown commonality in liquidity, adverse selection costs, and inventory carrying costs. We provide some evidence suggesting that market-wide divergence in opinions among traders is (at least partially) responsible, for this co-movement.

At a time when bull and bear swings in the market seem to have become a daily phenomenon and various trading strategies (machine and human) are being blamed for causing excessive volatility and possible market failures, we believe that it is very important to start taking a closer look at order-flow and order-flow variability. This study provides a lead in this direction. Further examination of these issues would be a reasonable direction for future research.

References

Admati, A. R., & Pfleiderer, P. (1988). A theory of intra-day patterns: volume and price variability. *Review of Financial Studies, 1,* 3-40.

Amihud, Y., & Haim, M. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics, 17,* 223-249.

Bessembinder, H., Chan, K., & Seguin, P. J. (1996). An empirical examination of information, differences of opinion, and trading activity. *Journal of Financial Economics, 40,* 105-34.

Brennan, M. J., Chordia, T., & Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics and the cross-sectional of expected returns. *Journal of Financial Economics, 49*, 345-373.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finace, 52,* 57-82.

Choie, K., & Hwang, S. J. (1994). Profitability of short-selling and exploitability of short information. *Journal of Portfolio Management, 20,* 33-38.

Chordia, T., Subrahmanyam, A., & Anshuman, V.R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics, 59,* 3-32.

Chordia, T., Richard, R., & Avanidhar, S. (2001). Market liquidity and trading activity. *Journal of Finance, 56,* 501-30.

Diether, K. B., MalLoy, C. J., & Scherbina, A. (2002). Difference of opinion and the cross section of stock returns. *Journal of Finance, 57*, 2113-2141.

Easley, D., & O'Hara, M. (1992). Time and the process of security price adjustment. *Journal of Finance, 47,* 577-605.

Fama, E. F., & French, K. R. (1989). Business conditions and Expected returns on Stocks and Bonds. *Journal of Financial Economics, 25*, 23-49.

Fama, E. F., & James, D. M. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy, 81*, 607-636.

Gallant, A., Rossi, P., & Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies,5,* 199-242.

Glosten, L. R., & Harris, L. E. (1988). Estimating the components of the bid ask spread. *Journal of Financial Economics, 21,* 123-142.

Grossman, S. J., & Merton, H. M. (1988). Liquidity and market structure. *Journal of Finance, 43,* 617-633.

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

Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance, 46,* 129-207.

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

Ho, T., & Stoll, H. (1983). The dynamics of dealer markets under competition. *Journal of Finance, 38,* 1053-1074.

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

Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis, 22,* 109-126.

Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica, 53,* 1315-1336.

Lee, C. M. C., & Ready, M. J. (1991). Inferring trade direction from intraday data. *Journal of Finance, 46,* 733-746.

Lee, C. M. C., & Radhakrishna, B. (2000). Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets, 3,* 83-111.

Lin, J. S., & Booth, G. (1995). Trade size and components of the bid-ask spreads. *Review of Financial Studies, 8,* 1153-1183.

Lo, A., & Wang, J. (2000). Trading volume: Definitions, data analysis, and implications of portfolio theory. *Review of Financial Studies, 13,* 257-300.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information.*The Journal of Finance, 42*, 483-510

Milgrom, P., & Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic Theory, 26,* 17-27

Miller, E. M. (1977). Risk, uncertainty and divergence of opinion. *Journal of Finance, 32*, 1151–1168.

Neal, R., & Wheatley, S. (1998). Adverse selection and bid-ask spreads: Evidence from closed-end funds. *Journal of Financial Markets, 1,* 121-149

Odders-White, E. R. (2000). On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets, 3,* 205-332.

Pontiff, J. (1996). Costly arbitrage: evidence from closed-end funds. *Quarterly Journal of Economics, 111,* 1135-1151.

Ritter, J. (1991). The long-run performance of initial public offerings. *Journal of Finance, 42,* 365-394.

Senchak, A. J., & Starks, L. T. (1993). Short-sale restrictions and market reaction to short-interest announcements. *Journal of Financial and Quantitative Analysis, 28,* 177-194.

<A level>Appendix

Table 1: Distribution of Firms Across the Sample Period

The count presents the number of firms in the given year for which all the data described in the table were available. All presented numbers are monthly arithmetic averages. SIGOF is the standard deviation of the 15-minute order imbalance. IA is the adverse selection cost component of the spread and INV is the inventory cost component of the spread (Lin, Sanger and Booth, 1988). Vol is the monthly traded volume, r is the risk-adjusted monthly stock return, S is the market capitalization of the firm and ANAL is the average number of analysts providing earnings estimate. DISP is the dispersion in analysts' forecasts.

Year	Count	SIGOF	IA	INV	VOL	r	MB	Size	ANAL	DISP
1993	1024	2.0449	0.3647	0.3588	38464.1231	0.0112	3.3509	2957022.82	13.3625	0.1452
1994	1176	2.0058	0.3954	0.3330	37890.9219	0.0023	2.9328	2682912.79	12.7174	0.1309
1995	1249	2.0850	0.3838	0.3360	41240.9121	0.0037	2.9418	2954618.12	12.1284	0.1332
1996	1321	2.2289	0.4000	0.3113	47536.4857	0.0081	3.0493	3431706.99	11.4832	0.1217
1997	1408	2.4401	0.4540	0.2361	56984.2854	0.0093	3.0276	4001075.08	10.6894	0.1113
1998	1398	2.6196	0.5069	0.1741	71378.6026	-0.0100	2.7896	4547333.39	10.2762	0.1088
1999	1359	2.8872	0.4835	0.1942	86380.2711	0.0008	2.4075	4907976.13	10.7654	0.1041
2000	1245	3.3258	0.4814	0.1866	112962.9103	0.0313	2.3686	5023442.19	10.6806	0.1041
2001	1162	4.6615	0.4593	0.1255	120635.3947	0.0215	2.2953	4755739.52	9.4567	0.1162
2002	1109	6.1435	0.4556	0.1217	125709.7064	0.0204	2.1486	4145072.55	8.1447	0.0888
2003	1081	7.3173	0.4247	0.1408	128425.5163	0.0167	2.0864	3624256.89	8.4979	0.1041
2004	1096	8.7262	0.4118	0.1084	134798.9144	0.0112	1.9763	3185072.07	7.2300	0.0965
2005	1064	10.0718	0.3944	0.0942	139945.1274	0.0067	1.8770	2704249.78	6.4440	0.0937

Table 2: Distribution of SIGOF by Industry

Industry	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Computer Manufacturing	3.794	3.705	4.335	4.560	4.910	4.762	6.329	8.050	9.628	11.044	10.941	11.851	12.507
Communication and electronic equipment	2.708	2.730	3.032	3.339	3.540	3.468	3.866	4.856	5.802	7.279	8.608	10.036	11.440
Oil and Gas	2.178	1.940	1.904	2.239	2.589	2.842	2.990	3.455	4.972	6.231	7.710	9.041	10.410
Computer and Data Processing Services	2.567	2.786	2.910	3.747	3.674	3.438	3.785	3.959	7.703	9.916	11.234	13.149	14.914
Optical, Medical, and Scientific instruments	2.476	2.171	2.443	2.978	2.942	2.794	3.072	4.088	5.965	7.344	8.193	9.396	10.510
Retailers	2.457	2.208	2.369	2.635	2.651	3.072	3.382	3.983	5.996	8.074	8.841	10.482	11.905
Wholesalers	1.981	1.915	1.994	2.092	2.135	2.477	2.757	3.097	4.634	6.640	8.419	10.350	12.242
Miscellaneous manufacturing	2.394	2.331	2.466	2.605	2.899	3.129	3.596	4.099	5.776	7.941	8.891	10.651	12.208
Health care and HMOs	2.209	2.288	2.440	2.639	2.659	2.698	2.880	3.889	6.139	8.079	9.191	10.855	12.382
Drugs and Genetic engineering	3.292	2.996	3.175	3.929	4.524	4.567	5.100	6.452	9.007	10.696	12.484	14.206	15.944
Miscellaneous Services	2.120	2.148	2.213	2.481	2.399	2.683	3.178	3.589	4.767	6.828	8.399	10.297	12.113
Transportation and Public utilities	1.996	2.040	2.038	2.171	2.353	2.744	3.210	4.000	6.073	7.810	9.166	10.776	12.322
Mining	1.998	1.998	2.045	2.147	2.226	2.574	2.884	2.954	4.783	5.844	6.164	6.979	7.670
Construction	1.996	1.937	1.779	2.112	2.036	2.136	2.469	2.852	5.044	7.160	8.353	10.161	11.816
Others	2.011	1.935	1.994	2.142	2.509	2.969	3.449	4.168	5.898	7.611	8.931	10.514	12.031

Table 3: Time-Series Andcross-Sectional Variation in Average SIGOF

Panel A: Cross-section divided into firm size (Market capitalization) quintiles:

Size Quintiles	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	1.466	1.438	1.466	1.504	1.541	1.545	1.569	1.611	2.079	2.673	3.126	3.673	4.197
2	1.709	1.601	1.646	1.729	1.836	1.922	1.919	2.051	3.003	4.363	5.329	6.558	7.721
3	1.816	1.779	1.832	1.926	2.079	2.24	2.319	2.595	3.831	5.425	6.732	8.23	9.681
4	2.192	2.171	2.312	2.468	2.766	3.085	3.345	3.948	6.018	8.305	9.889	11.942	13.877
5	3.489	3.495	3.764	4.377	5.216	5.862	7.032	8.129	11.306	14.355	16.478	19.218	21.804

Panel B: Cross-section divided into firm market-to-book ratio quintiles:

MB Quintiles	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	1.991	1.874	1.911	1.96	2.074	2.161	2.289	2.423	3.181	4.46	5.449	6.631	7.765
2	1.949	1.911	1.954	2.013	2.205	2.318	2.393	2.718	4.151	5.835	7.022	8.54	9.976
3	2.191	2.1	2.147	2.352	2.586	2.795	3.105	3.553	5.344	7.198	8.68	10.41	12.078
4	2.391	2.307	2.444	2.654	2.976	3.303	3.624	4.24	6.387	8.62	10.102	12.085	13.942
5	2.966	2.88	3.067	3.426	3.868	4.342	5.198	6.174	8.418	10.802	12.543	14.713	16.775

Panel C: Cross-section divided into quintiles of DISP (Dispersion in analyst forecast):

DISP Quintiles	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	1.56	1.493	1.534	1.604	1.659	1.741	1.73	1.948	2.758	3.988	4.767	5.847	6.851
2	1.753	1.66	1.709	1.777	1.928	2.095	2.1	2.45	3.472	5.319	6.42	8.018	9.492
3	2.005	2.003	2.155	2.221	2.322	2.632	2.826	3.408	5.033	7.337	8.782	10.799	12.674
4	2.563	2.54	2.686	2.945	3.176	3.624	4.21	5.178	7.665	10.32	12.415	14.883	17.258
5	3.812	3.753	3.991	4.576	5.332	6.145	6.972	8.295	11.614	14.408	16.79	19.447	22.035

Table 4: Distribution of SIGOF

Panel A: Distribution of SIGOF by market-to-book ratio andsize quintiles:

		Quintiles of MB				
		1	2	3	4	
Quintiles of size	1	1.916	1.897	1.953	2.022	2.799
	2	2.571	2.859	2.090	2.613	3.626
	3	3.105	3.362	3.182	3.304	4.648
	4	5.219	4.500	5.060	5.929	5.414
	5	8.056	7.500	7.431	7.709	7.253

Panel B: Distribution of SIGOF by market-to-book ratio and DISP (Dispersion in analysts' forecasts) quintiles:

	Quintiles of MB				
	1	2	3	4	5
1	2.517	2.702	2.785	2.911	3.148
2	3.632	3.538	3.563	3.902	4.196
3	4.451	4.539	4.714	4.691	5.892
4	4.456	4.516	4.500	4.882	5.825
5	5.932	6.398	7.088	8.075	8.298

Panel C: Distribution of SIGOF by size andDISP quintiles:

		Quintiles of Size				
		1	2	3	4	5
Quintiles of DISP	1	2.141	2.692	3.398	5.210	7.690
	2	2.968	3.608	4.258	5.553	8.804
	3	2.220	2.639	2.960	3.676	5.756
	4	3.172	4.582	4.566	5.450	8.064
	5	3.033	6.714	6.280	6.904	9.958

Table 5: Non-Parametric Correlation Coefficients (Spearman's Rank Correlation)

Time-series averages of the monthly cross-sectional correlation coefficients; SIGOF is the variability of 15-minute order-flow within a month; s_{Vol} is the standard deviation of 15-minute trading volume, within a month. DVIA is the per dollar adverse selection cost. It is calculated as the adverse selection cost component times the effective spread divided by the trading price. Vol represents the total trading volume within the give month. Ret is raw holding period return while risk-adjusted return is calculated using the four factor model described in Equation 3. MB is the market-to-book ratio of the firm, ANAL is the total number of analysts providing earnings forecasts for a given firm while DISP is the dispersion in their forecasts. Size is the market capitalization of the firm.

	SIGOF	σ_{Vol}	DVIA	Vol	ret	ret (risk adjusted)	MB	ANAL	DISP
SIGOF	1								
σ_{Vol}	0.488	1							
DVIA	-0.572	-0.376	1						
Vol	0.660	0.849	-0.614	1					
ret	0.027	0.021	-0.034	0.013	1				
ret (risk Adj.)	0.008	0.003	-0.020	-0.011	0.972	1			
MB	0.342	0.014	-0.335	0.269	0.102	0.091	1		
ANAL	0.713	0.247	-0.672	0.757	-0.002	-0.019	0.249	1	
DISP	0.036	0.011	-0.054	0.032	-0.016	-0.018	-0.177	0.066	1
Size	0.782	0.396	-0.798	0.814	0.016	-0.004	0.386	0.784	-0.003

Table 6: Attributing SIGOF to firm Andtrading Characteristics

$$SIGOF_{i,t} = \alpha_t + \beta_{1,t} \times \ln\left(\sigma_{VOL,i,t}\right) + \beta_{2,t} \times \ln\left(Vol_{i,t}\right) + \beta_{3,t} \times \ln\left(Size_{i,t}\right) + \beta_{4,t} \times IA_{i,t} +$$
$$+ \beta_{5,t} \times INV_{i,t} + \beta_{6,t} \times \ln\left(Anal_{i,t}\right) + \beta_{7,t} \times DISP_{i,t} + \beta_{8,t} \times MB_{i,t} + \varepsilon_{i,t}$$

The cross sectional model is estimated in each month of the sample period (January 1993 to December 2005). The table reports the time-series averages of the slope coefficients. The t-statistics corresponds to the test $Average(\beta) = 0$. The table also reports the average adjusted R^2. The explanatory variables are, dispersion in analysts' forecasts (DISP), market to book ratio of the firm (MB) andnatural logarithm of: trading volume (Vol), 15-minute variability in trading volume within the month (s_{Vol}), andnumber of analysts' providing earnings forecasts (ANAL). IA andINV are the adverse selection andinventory cost component of the spread respectively. The constant term is not reported.

$\ln(\sigma_{Vol})$	ln(Vol)	ln(Size)	IA	INV	ln(Anal)	DISP	MB	Mean Adj R^2
1.473								0.475
(20.62)								
	1.103							0.551
	(20.32)							
		1.375						0.571
		(20.72)						
			-0.324					0.032
			(-3.87)					
				-0.905				0.016
				(-4.08)				
0.607	0.635	0.809						0.638
(20.94)	(12.36)	(25.18)						
		1.371			0.853	0.506	-0.072	0.612
		(23.16)			(11.15)	(14.11)	(-4.27)	
0.263	0.703	0.617	-2.673		0.201	0.467	-0.104	0.681
(11.32)	(14.84)	(25.93)	(-7.65)		(4.27)	(19.14)	(-3.09)	
0.197	0.851	0.619		-0.509	0.182	0.421	-0.015	0.670
(10.11)	(15.14)	(26.63)		(-3.12)	(4.22)	(16.57)	(-3.22)	
0.132	0.811	0.622	-2.147	-1.381	0.175	0.517	-0.017	0.699
(9.1)	(14.53)	(27.41)	(-10.55)	(-5.47)	(4.15)	(18.18)	(-3.42)	

*t-statistics are given in the parenthesis bellow the coefficients.

Table 7: Attributing Time-Series Changes in SIGOF to Changes in Systematic Andfirm Specific Factors

$$DSIGOF_{i,t} = \alpha_i + \beta_{1,i} \times DNOIC_{i,t} + \beta_{2,i} \times DVol_{i,t} + \beta_{3,i} \times DSigVol_{i,t} + \beta_{4,i} \times r_{i,t} + \beta_{5,i} \times DANAL_{i,t} + \beta_{6,i} \times DDISP_{i,t} + \beta_{7,i} \times DMB_{i,t} + \varepsilon_{i,t}$$

Monthly proportional change in the individual stock's order-flow variability (SIGOF) is regressed in time-series on the cotemporaneous explanatory variables. The table reports the cross sectional averages of the coefficients. The t-statistics corresponds to the test $Average(\beta) = 0$. The table also reports the cross sectional average adjusted R^2. The explanatory variables are, risk adjusted return andmonthly proportional change in: number of S&P open interest contracts (NOIC), trading volume (Vol), monthly volatility of the trading volume (s_{Vol}), the number of analysts' providing earnings forecasts (ANAL), dispersion in analysts' forecasts (DISP), andmarket-to-book ratio of the firm (MB).

DNOIC	DVol	DSigVol	beta adj ret (VW)	DANAL	DDISP	DMB	Mean Adj R2
0.012							0.019
(1.86)							
	0.108						0.321
	(49.85)						
		0.081					0.272
		(47.61)					
			0.318				0.047
			(21.05)				
				-0.009			0.011
				(-0.87)			
					-0.011		0.017
					(2.17)		
		0.074	0.271		-0.021	0.03	0.255
		(26.53)	(2.15)		(-0.98)	(0.16)	
		0.072	0.323	-0.022	-0.032	-0.01	0.259
		(21.33)	(3.26)	(-0.87)	(-0.83)	(-0.79)	
0.027		0.065	0.258	-0.045	-0.033	-0.061	0.261
(0.65)		(19.75)	(3.49)	(-1.07)	(-0.86)	(-0.81)	

*t-statistics are given in the parenthesis bellow the coefficients.

Table 8: Market-Wide Commonality in SIGOF Andliquidity

Monthly proportional change in individual stock's order-flow variability (SIGOF) is regressed in time-series on proportional change in the equal-weighted average order-flow variability for all stocks in the sample (the 'market').

$$DSIGOF_{i,t} = \alpha_i + \beta_i \times DSIGOF_{M,t} + \gamma_{i,1} r_{m,t-1} + \gamma_{i,2} r_{m,t} + \gamma_{i,3} r_{m,t+1} + \gamma_{i,4} \ln\left(\frac{P_{i,t,\max}}{P_{i,t,\min}}\right) + \varepsilon_{i,t}$$

The right handside control variables include a lead anda lag market return $\left(r_{m,t+1} \text{ and } r_{m,t-1}\right)$, anda measure of monthly volatility (Natural logarithm of the ratio of the maximum stock price to the minimum stock price in the given month).

The procedure is repeated for two liquidity measures: QSPR (the quoted spread) andPQSPR (the proportional quoted spread), proportional change in monthly adverse selection cost of trading (DDVIA) andthe monthly proportional change in inventory cost incurred by the market maker (DDVINV).

The letter D denotes proportional change. Therefore for measure M, $DM_t = \left(M_t - M_{t-1}\right)/M_{t-1}$

The table below presents the statistics for the beta (β_i) coefficients from the above equation (equation 7). '%Positive' reports the percentage of positive beta coefficients, while '% + Sig' gives the percentage significant at the 5% one-tailed critical value.

	DQSPR	DPQSPR	DSIGOF	DDVIA	DDVINV
Adj R2 Mean	12.10%	15.87%	6.04%	6.41%	1.87%
Adj R2 Median	8.01%	13.69%	4.32%	4.01%	0.13%
% Positive	92.72%	96.22%	87.34%	93.11%	77.01%
% + Sig.	66.91%	77.96%	49.21%	57.83%	26.87%
% Negative	7.28%	3.78%	12.66%	6.89%	22.99%
% - Sig.	0.55%	0.21%	1.21%	0.41%	2.74%

Table 9: Market-Wide Commonality in Levels of Liquidity

(As measured by Quoted spread (QSPR) andproportional quoted spread (PQSPR)), adverse selection cost per dollar of trade (DVIA) andInventory cost per dollar of trade (DVINV):

Monthly levels of individual stock's order-flow variability (SIGOF) are regressed in time-series on the equal-weighted average order-flow variability for all stocks in the sample (the 'market').

$$SIGOF_{i,t} = \alpha_i + \beta_i \times SIGOF_{M,t} + \gamma_{i,1} r_{m,t-1} + \gamma_{i,2} r_{m,t} + \gamma_{i,3} r_{m,t+1} + \gamma_{i,4} \ln\left(\frac{P_{i,t,\max}}{P_{i,t,\min}}\right) + \varepsilon_{i,t}$$

The right hand side control variables include a lead anda lag market return $\left(r_{m,t+1} \text{ and } r_{m,t-1}\right)$, anda measure of monthly volatility (Natural logarithm of the ratio of the maximum stock price to the minimum stock price in the given month).

The procedure is repeated for two liquidity measures: QSPR (the quoted spread) andPQSPR (the proportional quoted spread), monthly adverse selection cost of trading (DVIA) andthe monthly inventory cost incurred by the market maker (DVINV).

Panel Apresents the statistics for the beta coefficients from the above equation. '%Positive' reports the percentage of positive beta coefficients, while '% + Sig' gives the percentage significant at the 5% one-tailed critical value.

Panel B repeats the analysis using the residuals from:

$$M_{i,t} = \alpha_{1,i} + \alpha_{2,i} SIGOF_{i,t} + \varepsilon_{i,t}$$

Where $M_{i,t}$ is a general representation for the quoted spread (QSPR), proportional quoted spread (PQSPR), adverse selection cost per dollar traded (DVIA) andthe inventory cost per dollar traded (DVINV), for firm i in month t.

Panel A: Commonality in Iiquidity, adverse selection cost and Inventory cost

	QSPR	PQSPR	SIGOF	DVIA	DVINV
Adj R2 Mean	63.44%	49.16%	46.60%	31.51%	34.30%
Adj R2 Median	74.29%	56.63%	49.63%	27.44%	30.75%
% Positive	95.65	91.47	85.68	91.18	87.33
% + Sig.	90.52	84.50	74.44	79.02	74.90

Panel B: Liquidity comovement, controlling for SIGOF

	QSPR	PQSPR	SIGOF	DVIA	DVINV
Adj R2 Mean	27.26%	24.20%		14.79%	15.40%
Adj R2 Median	26.56%	23.41%		11.08%	12.57%
% Positive	91.93	91.35		91.53	83.22
% + Sig.	81.86	78.80		75.53	63.88

Table 10: The Contribution of Co-Movement in SIGOF to Co-Movement In Liquidity

Panel A presents the mean andthe median pair-wise correlation, run across all 5418 firms in the sample. All pairs with less than 20 observations are omitted from the analysis.

Panel B presents the pair-wise correlation between the same set of variables, controlling for the effect of SIGOF. We regress the average monthly liquidity measures on contemporaneous SIGOF for the stock andexamine the cross stock correlation of the residuals $\left(\varepsilon_{i,t}\right)$ from the following model: $M_{i,t} = \alpha_{1,i} + \alpha_{2,i} SIGOF_{i,t} + \varepsilon_{i,t}$

Where $M_{i,t}$ is a general representation for the monthly average quoted spread (QSPR), proportional quoted spread (PQSPR), adverse selection cost per dollar traded (DVIA) andthe inventory cost per dollar traded (DVINV), for firm i in month t.

This analysis helps to identify the contribution of co-movement in SIGOF to co-movement I liquidity.

Panel A

	QSPR	PQSPR	SIGOF	DVIA	DVINV
Mean Correlation	0.5413	0.4620	0.4236	0.2064	0.2594
Median Correlation	0.6029	0.4373	0.4758	0.2250	0.2173
Number of Observations	13,422,532	13,422,532	13,422,532	13,422,532	13,422,532

Panel B

	QSPR	PQSPR	SIGOF	DVIA	DVINV
Mean Correlation	0.2200	0.1190		0.0586	0.1721
Median Correlation	0.2720	0.2815		0.0374	0.0695

Table 11: Some Explanations for Market Wide Commonality in SIGOF

Monthly individual stock's order-flow variability (SIGOF) is regressed in time-series on a set of lagged control variables. The set of control variables are: SIGOF, volume volatility, number of S&P open interest contracts (NOIC), value weighted market return, market capitalization of the firm (Size), trading volume (Vol), number of analysts providing earnings' forecasts for the firm (ANAL), Market to Book ratio (MB), anddispersion in analysts' forecasts (DISP).

We use the residuals $\left(\varepsilon_{i,t}\right)$ from

$$SIGOF_{i,t} = \alpha_{1,i} + \sum_{j} \alpha_{2,i,j} \left(Control\ Variable_j\right)_{i,t} + \varepsilon_{i,t}$$

to run co-movement analysis, using the equation:

$$\varepsilon_{i,t} = \alpha_i + \beta_i \times \varepsilon_{M,t} + \gamma_{i,1} r_{m,t+1} + \gamma_{i,2} r_{m,t-1} + \gamma_{i,3} \ln\left(\frac{P_{i,t,\max}}{P_{i,t,\min}}\right) + \psi_{i,t}$$

The table presents, the cross-sectional average beta (β_i), andthe corresponding t-statistics. '%Positive' reports the percentage of positive beta coefficients, while '% + Sig' gives the percentage significant at the 5% one-tailed critical value.

	No Controls	SIGOF and SIGVOL	SIGOF, ln(NOIC), market ret.	SIGOF, ln(Size), Ln(Vol), ln(ANAL), MB, DISP, SIGVOL	SIGOF, ln(Size), Ln(Vol), ln(ANAL), MB, DISP, SIGVOL, Ln(NOIC), market ret.
	SIGOF	SIGOF	SIGOF	SIGOF	SIGOF
Mean beta	0.834	0.844	0.762	0.637	0.244
t-stat	35.538	37.202	37.083	26.013	7.403
Adj R2 Mean	46.60%	6.43%	6.64%	2.12%	0.60%
Adj R2 Median	49.63%	4.37%	3.37%	1.46%	-0.42%
% Positive	85.675	84.364	85.045	72.454	70.951
% + Sig.	74.44	57.32	55.55	19.53	16.41