

Institutional Trading and Spread

Malay K. Dey

Morgan State University

B. Radhakrishna

UBS Warburg

[This Draft: March 2003]

This paper was previously circulated as “Institutional Trading, Trading Volume, and Spread”. We are grateful to NYSE for providing us with the TORQ data and to Murat Aydogdu, Ben Branch, Shanta Hegde, Jennifer Koski, Robert Miller, Larry Wall, C. J. Wang, and participants at Finance workshop at Concordia University, NBER Market Microstructure Group meeting, and Washington Area Finance Conference for their comments and numerous helpful suggestions. We are solely responsible for any error. The views expressed in this paper should not be interpreted as those of UBS Warburg. Comments are welcome.

Please address correspondence to:

Malay K. Dey, Dept. of Accounting and Finance, McMechen Hall 613, Morgan State University, 1700 E. Cold Spring Lane, Baltimore, MD 21251, Tel. (443) 885-1694, Email: mdey@morgan.edu.

Institutional Trading and Spread

Abstract

The effect of institutional trading on bid-ask spread is of interest to academics, regulators, and market makers. However, there is no direct empirical evidence on the relation between institutional trading and spread. In this paper, we present evidence from transactions data that in addition to the effect of large size and high trading frequency associated with institutional trades, institutional trades *per se* affect spreads. The inverse relationship between spread and institutional trading is non linear and the precise nature of the relation depends on the level of institutional trading, and the level of activity (actively traded vs. less active) of the underlying stock. The inverse relation (downward slope) between institutional trading and percentage spread becomes steeper before it flattens out (decreases) at the highest level of institutional trading. Further, percentage spread for actively traded stocks are initially higher than that of their less active counterparts, but the percentage spread decreases (increases) for active (less active) stocks as institutional trading increases. We find evidence that institutional sells increase the adverse selection component, but large and frequent trading by institutions reduces the order processing costs, which outweighs the corresponding increase in adverse selection costs.

Keywords: Institutional Trading, Spread decomposition, information, liquidity.

JEL Classification: G19.

Institutional Trading and Spread

I. Introduction

Since Demsetz (1968) bid-ask spread is recognized as the price of liquidity provided by the dealers in a security market. Schwartz (1988) identifies four factors, namely, activity, risk, information, and competition as determinants of spread. Existing literature¹ finds the following: trade size, trading volume, and number of trades measure activity and bear an inverse relationship with spread; risk measured by volatility, market power (absence of competition) in the dealership market, and stock price, to the extent it proxies for information risk increase spread. Several recent studies report ownership structure as a proxy for information and thus a determinant of spread. For example, Chung and Charoenwong (1998), Heflin and Shaw (2000), and Sarin, Shastri, and Shastri (2000) find insider trading, block holding, and institutional holding respectively increase spread, while Kini and Mian (1995) report block holding does not affect, but institutional holding decreases spread. Yet, there is no direct evidence on how institutional *trading*² affects spread.

In this paper, we examine for the first time using transactions data (TORQ) how institutional trading affects spread. The role of institutional trading in the determination

¹ Existing literature on determinants of bid-ask spread in equity markets is quite extensive. The following are a few that are directly related to our study - Branch and Freed (1977), McNish and Wood (1992), Laux (1993, 1995), Kini and Mian (1995), Klock and McCormick (2000), Harris (1994), and Heflin and Shaw (2000).

² Institutional trading is a flow while institutional holding is a stock variable. An institution may hold a certain quantity of a particular stock at the beginning of a period, sell and buy the same quantity during a given period, and ends up with no change in inventory at the end of the period. Thus they have different distributional properties. Chung and Charoenwong (1998) find insider trading increases spread in a cross sectional regression, but find no changes in spread on insider trading days. It also remains to be seen with ample data being available, whether institutional holding and institutional trading both have similar information content and thus one of those loses its explanatory power in a combined model.

of spread is interesting since institutions have resources and incentives to collect, analyze, and trade on the basis of costly information, and thus their trades may contain value related information. Sias, Starks, and Titman (2001) provide a comprehensive summary of the different hypotheses behind information content in institutional trades. This information content in institutional trade should increase the risk of information trading and thus the spread. On the contrary, Scholes (1972) contend that institutional trades are motivated by both information and liquidity. If institutional trades provide liquidity (non-information or non-value related information), then the increased trading activity due to institutional trades should decrease spread.

The discovery of information content in institutional trades is complicated by the fact that exchange rules prohibit disclosure of trader identification and thus mostly it is only through a reasonable proxy that the information content in institutional trades is tested. Trade size is a commonly used proxy to identify institutional trades, since institutions generally trade in large sizes³. However, there are two problems with that approach. First, the effect of trade size, particularly in the context of setting spread, is not conclusive. McNish and Wood (1992) discuss the trading cost literature that argues large trading volume and frequent trading reduce spread. On the other hand, confirming information asymmetry hypothesis that large trades contain value-related information, Hasbrouck (1991) finds large trades cause spreads to widen, and Lin, Sanger and Booth – hereafter LSB (1995) report an increasing, but non- monotonic relation between effective spread and trade size. Second, it is not clear whether the information content in

³ Lee and Radhakrishna (2000) confirm earlier evidence that institutions tend to trade in larger volumes.

institutional trades is due to the large size of institutional trades, or there is information content in institutional trades *per se*.

The TORQ data enables us to identify institutional trading⁴ and thus we test directly and not through a size proxy how institutional trading affects spread. This separation between institutional and large trades is important since based on recent evidence from Heflin and Shaw (2000) and Chakrovarty (2001), we argue that institutional trades *per se* irrespective of their size affect spread, the price of liquidity. Heflin and Shaw (2000) find not only percentage held by block holders, but also the percentage of non-block holding insiders affect spread; Chakrovarty (2001) identifies institutional trades from TORQ data and reports that a majority of institutional trades are not large trades, but mid-size trades that capture most of the price effect due to trade size. We contend that the effect of intra-day institutional trading on spread goes beyond its value related information content and may indicate other non-value related information like trading cost and liquidity motive.

We test directly how institutional trading affects spread after controlling for effects on spread due to high trading volume, and frequent trading usually associated with institutional trading. Most studies on the determinants of spread focus on the supply side⁵ of dealership market i.e., competition in the dealer market, and use a cross sectional regression approach. Our approach is different from previous studies in three important ways. First, we focus on the demand side⁶ (investors' characteristics) to determine the

⁴ TORQ does not identify institutional trades directly, but Radhakrishna (1995), and Chakrovarty (2001) layout a classification scheme to identify institutional trades. The scheme is discussed in the data section.

⁵ In a securities market, dealers supply while investors demand liquidity. Structure of the dealers' market represents the supply side, while traders' preferences such as trading frequency, and order size represent the demand side of the market for liquidity services.

⁶ Recent studies on the demand side focus on the ownership structure e.g., block holding (Heflin and Shaw [2000]), and insider holding (Chung and Charoenwong [1998]) as a determinant of spread.

relation between spread and institutional trading; second, we use transactions data that enables us to identify intra-day institutional trading compared to institutional holding; and third, we use a panel data approach accounting specifically for both serial and contemporaneous correlation in the error terms. The panel data estimates use a two stage GLS method and thus provide improvements over OLS estimates in terms of efficiency and bias (Kennedy [1998]).

Further, we decompose spread into two different components, adverse selection, and order processing⁷ costs and investigate in a simultaneous equations framework how these components vary with changes in trading volume, net order flow (buy-sell), and institutional trading. We assume contemporaneous correlation between the disturbances of the two equations and use an SUR (Seemingly Unrelated Regression) analysis to find the significant determinants of the adverse selection and the order processing components of the spread for our sample firms.

Our results show that institutional trading explains a significant portion of the variation in percentage spread across all specified multivariate models. Except in one model mentioned later, institutional trading proportion is consistently inversely related to percentage spreads. We find that this negative slope is provided by both institutional buys and sells alike. We also find that the negative slope is not constant; it becomes steeper before it flattens out (decreases) at the highest level of institutional trading. The non-linear relationship between institutional trading and percentage spread becomes more apparent as we distinguish between active and less active stocks. We find that in the

⁷ Arguably there is also inventory effect in spread. Information asymmetry models acknowledge but nevertheless, ignore inventory effect in spread. In this paper, we use a decomposition technique based on Stoll (1989) that ignores inventory effect.

beginning, percentage spread for active stocks is higher than that for less active stocks, but as institutional trading increases, percentage spread decreases (increases) for active (less active) stocks. The overall inverse relation between institutional trading and percentage spread in our sample is due to the high percentage of active stocks in our sample that exhibit inverse relation between percentage spread and institutional trading.

Results from a SUR analysis show that the adverse selection costs tend to increase while order processing costs tend to decrease with increases in institutional trades. Further institutional buys do not provide value related information, but institutional sells do. This is consistent with the conjecture that institutional buys follow security analysts' buy recommendations that is already public; however institutional sells provide new information, since sell recommendations are extremely rare (Barber et al [2001]).

The rest of this paper is organized as follows. In section II, we present our hypotheses, describe an empirical model for spread using panel data regressions and the data, and explain the results. In section III, we describe our method of decomposing the spread into order processing and adverse selection costs components based on Stoll (1989) and report results from a SUR analysis of the determinants of those components. Section IV concludes the paper.

II. Hypothesis, Regression Models, Data Description, and Results

II.A. Developing Hypotheses

In a multivariate panel data regression framework, we test the hypothesis that institutional trades *per se* affect spread, the price of liquidity. Institutions trade in large

sizes, and trade frequently. Thus if institutional trades *per se* have no explanatory power beyond those revealed by trade size, and trading frequency then we would expect in a multivariate regression with proxies for trade size and frequency, institutional trading to have no effect on spread. On the contrary, we hypothesize that after adjusting for documented effects due to trading volume, number of trades, price, and volatility, the variation in bid-ask spread can be explained by institutional trading. Since our interest is in estimating the information content in the independent variables revealed in spread, we use abnormal i.e., unanticipated trading volume (daily trading volume over its time series average), and abnormal i.e., unanticipated trading (daily number of trades over its time series average) instead of levels in our empirical model. Further, prompted by the findings in Koski and Michaely (2000) that buys and sells provide different information for different trade sizes, we determine if the effect of institutional trading *per se* on percentage spread depends on the direction of an institutional trade - buy or sell. Stated in alternate form, we hypothesize:

H1: The bid-ask spread should vary significantly with institutional trading after controlling for abnormal trading (number of trades), abnormal trading volume, price, and volatility. Further, institutional buys and sells will have a differential effect on the bid-ask spread after controlling for abnormal trading (number of trades), abnormal trading volume, price, and volatility.

II.B. Panel Data Regression Model

We use a multivariate, panel data (time series – cross sectional) regression framework to investigate the effect of institutional trading on bid-ask spread. Most studies on the determinants of spread use pooled OLS estimates of the parameters of a

regression model. OLS estimates ignore the covariance structure of the error term both across firms and over time, and may also induce simultaneous equations bias, if the regressor variables are not truly exogenous. In such instances, panel data estimates using two stage GLS methods result in more efficient, unbiased estimates.

We propose a random effect model; specifically, the random disturbances are both serially and contemporaneously correlated. Thus we consider an $AR(1)$ process with contemporaneous correlation for the disturbance term. In our model for the spread, the serial correlation may be due to lagged spread or lagged values of the independent variables or their interactions. Kim and Ogden (1996) find higher order serial correlation for the spread, and Peles (1992) report contemporaneous correlation over time among equity trading of institutional investors. Parks (1967) provide consistent and efficient estimates of the parameters when disturbances follow a first order auto regressive process - $AR(1)$ with contemporaneous correlation.

The following is the basic regression model for our panel data:

$$Spr_{nt} = \alpha + \beta_1 NtrPct_{nt} + \beta_2 DlyAv_{nt} + \beta_3 Pr_{nt} + \beta_4 Volty_{nt} + \beta_5 Instpr_{nt} + \varepsilon_{nt} \quad \dots(1)$$

where:

$n = 1 \dots N$; number of stocks⁸ in sample, $t = 1 \dots T$; number of trading days

Spr_{nt} = Effective or percentage spread for the nth stock on day t

$NtrPct_{nt}$ = Number of trades for stock n on day t expressed as a proportion of average number of daily trades over the entire sample period

$DlyAv_{nt}$ = Volume of trade for stock n on day t expressed as a proportion of average daily trading volume over the entire sample period

Pr_{nt} = Closing price for stock n on day t

$Volty_{nt}$ = Realized volatility computed as standard deviation of prices for stock n on day t

⁸ We use the words firm and stock interchangeably.

$Instpr_{nt}$ = Institutional trading (number of trades) as a proportion of total trading for stock n on day t .

Further, the error structure is assumed as follows:

$$E(\varepsilon_{nt}^2) = \sigma_{nn},$$

$$E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij} \text{ where } i \in n, \text{ and } j \in n \text{ (due to contemporaneous correlation),}$$

$$\varepsilon_{nt} = \rho_n \varepsilon_{n,t-1} + \mu_{nt} \text{ (disturbance term is } AR(1)),$$

$$\mu_{nt} \sim N(0, \phi_{nn}),$$

$$E(\varepsilon_{i,t-1}\mu_{jt}) = 0,$$

$$E(\mu_{it}\mu_{js}) = 0 \text{ where } t \neq s \text{ (cross correlation is zero),}$$

$$E(\mu_{it}\mu_{jt}) = \phi_{ij} \text{ (contemporaneous correlation).}$$

The panel data model with autocorrelated error recognizes and resolves an apparent endogeneity problem in an OLS regression model i.e., a significant relation between bid ask spread and institutional trading may suggest institutional trades are motivated by low bid ask spread (Bertsimas and Lo [1998]⁹) and thus institutional trading is not truly exogenous. This simultaneous equations bias of OLS estimates stems from the condition that $E(Spr_{nt}\varepsilon_{nt}) \neq 0$. A two stage least squares based on instrumental variable or correlated disturbance as in SUR approach resolves the problem. Note that in the above model $\varepsilon_{nt} = \rho_n \varepsilon_{n,t-1} + \mu_{nt}$ and thus equation (1) can be written as follows:

$$Spr_{nt} = \alpha + \beta_1 NtrPct_{nt} + \beta_2 DlyAv_{nt} + \beta_3 Pr_{nt} + \beta_4 Volty_{nt} + \beta_5 Instpr_{nt} + \hat{\rho}\varepsilon_{nt-1} + \mu_{nt}$$

where $\hat{\rho}$ is a consistent and efficient estimator of ρ , and $E(Spr_{nt}\mu_{nt}) = 0$. The

parameters of the model are two stage, GLS estimates and unbiased.

We further break up institutional trading into institutional buy and institutional sell and rerun regression model (1) by replacing institutional trading with institutional

⁹ Bertsimas and Lo (1998) propose a dynamic equilibrium model in which portfolio managers minimize execution costs of trading. To the extent execution costs represent order processing and similar costs, those form an integral component of spread.

buy and sell to investigate if institutional trading *per se* or the direction of institutional trading has any effect on the spread. We conduct an F-test to test the null hypothesis that the coefficients of buy and sell are similar. We also categorize institutional trading into high, low, and medium classes to gain further insight into the nature of the relation between institutional trading and spread. Specifically, we investigate if the relation (slope) between institutional trading and spread changes over levels (high, medium, or low) of institutional trading.

Finally, we use a number of trades dummy (the dummy variable attains a value of 1 if average number of daily trades for a stock is greater than 45, the median of average number of daily trades for the sample stocks; 0 otherwise) to separate active trading stocks from less active trading stocks in the regression. We create an interaction variable by multiplying the dummy variable with institutional trading proportion such that for less active stocks the interaction variable assumes a value zero, while for active stocks the interaction variable retains its value as the proportion of institutional trading. Used in conjunction with the number of trade dummy variable, the interaction variable will shed light on how institutional trading affects spread differently between active and less active stocks.

II.C. Data

To estimate the parameters of our regression model, we use a unique data set (TORQ) that allows us to identify the order origination for a trade as institutional or otherwise. The TORQ files released by NYSE were prepared under the supervision of Professor Joel Hasbrouck during his tenure as a Visiting Economist to the NYSE. TORQ

contains data on trades, quotes, and orders, and audit trail data for a sample of 144 NYSE stocks for the three months (63 trading days) from November 1990 through January 1991¹⁰. These stocks represent a size stratified random sample of stocks in the NYSE and thus cover the broad spectrum of NYSE stocks.

Koski and Scruggs (1998) note that the marginal contribution of TORQ data over ISSM or TAQ data is in providing identification for traders' classes, as agencies and proprietary, individuals, and dealers. The TORQ classification identifies the following three traders classes based on member firm's disclosure about a trade. *P* is a trade from a member for its personal account, not including specialists' trades, *I* is for an individual investor, *A* for an agency trade where the member broker acts as an agent for other traders, mostly institutions. Lee and Radhakrishna (2000) and Radhakrishna (1995) identify agency trades (classified as *A*) as initiated by institutions. Chakrovarty (2001) classifies agency (classification *A*) and proprietary and inter member trades (classification *P*) as institutional and relates the price effect of such trades to price effect of mid size, and large trades. Both Radhakrishna (1995) and Chakrovarty (2001) report that although trades classified under *A* may have included inter member (member firms) trades, their "discussions with exchange officials indicate that a vast majority of *A* trades are non-exchange member trades".

We impose a restriction that is common among studies that study effective spread or the components of the spread to ensure adequacy of data in estimation. We select all firms in the data set that have on average 20 trades per day or more during the sample period. Further, in classifying the trades, whenever there are executions of multiple

¹⁰ For more details on the TORQ database and on trading procedures at the NYSE, see Hasbrouck (1992).

orders on the active side of a trade, we take the trader class of the largest order on the active side as the initiator of that trade. The active side of the trade is determined using the Lee-Ready (1991) algorithm. This reduces our sample to 65 firms.

Of the 65 firms that survived our initial cutoff, 14 firms have one or more days of missing observations due to reasons like trading halts. We leave out the firms with trading halt days from our study since the effect of trading halts on spreads and the price discovery process is unique and beyond the realm of this paper. Thus we have a balanced panel of 51 firms with 63 days data that we use in our panel regression.¹¹

The panel data includes daily data on effective and percentage spreads, proportion of institutional trading, abnormal trading (number of trades), abnormal trading volume, stock price, and volatility for the 51 firms in our sample. The dependent variable is percentage spread computed as effective spread divided by the prevailing mid-quote. For each firm, we calculate effective bid-ask spread for each trade as twice the absolute value of the difference between trade price and the prevailing mid-quote i.e., $[Abs (price-mid\ quote)*2]$. The mean effective spread is the average of all effective spreads across all trades in a day and the mean percentage spread is the average of *(effective spread/prevailing mid-quote)* across all trades in a day.

We determine institutional trading as the proportion of trades by institutions in a stock, for each day. We also compute institutional buy and sell proportions for each stock day as the number of institutional buy (sell) over total number of buy (sell). The

¹¹ We explored another alternative i.e., to consider all the firms (65) but for fewer days (61). However, there was a problem in deciding which 2 days to delete for firms that have full 63 days data. Since any choice on this was likely to be arbitrary, we chose to use a balanced panel of 51 firms over 63 days.

“number of trades” variable is computed as the number of trades each day divided by the average number of trades per day for the stock over the sample period. Thus this is a measure of abnormal trading in each day. Trading above (below) mean would give this variable a value higher (lower) than one. The “trading volume” variable is computed by dividing the daily share volume for the stock by the average daily share volume over the sample period. Therefore, this variable also measures abnormal trading volume and has values above (below) one when trading volume is higher (lower) than the average daily share volume. The price variable is the closing price of the stock.¹²

In Table 1, we present means of the computed statistics of the variables used in our analysis for our entire sample of 65 firms. Thus we report the means of the cross sections of stock means, medians, and standard deviations. The inter-firm mean (median) spread is .126 (.12) that is about an eighth, and the inter-firm mean standard deviation is quite low at .02. The largest spread in our sample is .514, approximately one-half, and the lowest .019 or approximately one-fiftieth. The inter-firm mean (and median) institutional trading in our sample is around 30%. Although institutions generally trade on a regular basis, on some trading days there is no institutional trading. Of the 4,075 (51 firms for 63 trading days, and 14 firms with 62 or 61 trading days) stock trading days covered in our sample there were 10 firm-days when there was no institutional trading, and one firm-day when all the trading was initiated by institutions. There is also sufficient variability in the number of trades and volume. The low and high trades (volumes) in our sample were 9% (3%) and 392% (1235%) respectively of average daily trades (volume).

¹² Using average price or price inverse over a day instead of closing price does not change the results, neither does volatility when measured as standard deviation of return instead of prices.

Finally, price and volatility measure information and stand alone risk associated with a stock. The average price of a stock in the sample is around 24\$, and mean volatility is around 3%. Unlike individuals, institutions generally hold expensive, less volatile, seasoned stocks (Barber and Odean [2000]). Thus our sample of stocks with prices and volatility ranging from \$2.50 to \$83.50 and 0.00 percent to 172 percent respectively appeal to a broad cross section of institutional and non-institutional investors as evidenced by daily institutional trading proportion ranging from 0.00 percent to 100 percent.

In Table 2A, we present pooled Pearson correlations computed from 3,213 (51*63) observations (firm days) for the dependent and independent variables in the model. In general, Panel A confirms some intuitively obvious relations, for example, a significantly (at greater than 90% level of confidence) positive correlation between effective and relative spread, and between abnormal trading and abnormal trading volume. Since percentage spread is computed as effective spread over mid quote, as expected price and volatility show an exactly opposite relation vis a vis effective (+ve) and percentage (-ve) spreads. Abnormal trading and abnormal trading volume maintain a highly significant (greater than 99% level of confidence) positive correlation (.5164), but are quite different in their relation with effective and percentage spreads. Regarding percentage spread, both abnormal trading and abnormal trading volume show insignificant relations; for effective spread the correlation coefficient switches from significantly positive (.0297) to significantly negative (-.0384). The correlation coefficient between institutional trading proportion and effective (percentage) spread is insignificant (significant). The pooled correlation coefficients in Panel A have to be

cautiously interpreted since the coefficients are estimated allowing for both time and cross sectional variation in the variables. A wide cross sectional variance might have influenced several of those coefficients.

We disallow the cross sectional variance and compute the means of time series correlations for each firm. In Panel B, we report those means along with the number of firms showing significant correlation. For all pair wise correlations, except between effective spread and abnormal trading, the means of stock correlation in Panel B are visually different from the pooled correlation coefficients reported in Panel A. For example, the pooled sample shows a statistically insignificant relation ($\hat{\rho} = -.0005$ with p-value .979) between effective spread and institutional trading in Panel A, whereas the mean of stock correlation (Panel B) is $-.095$ with 13 firms showing a significant relation between institutional trading and effective spread. On the contrary, the relation between percentage spread and institutional trading is significant at less than one percent level ($\hat{\rho} = -.2617$ with p value $<.0001$) in the pooled sample, but the mean of stock correlation is $-.0593$ with almost similar number of firms (14) having a significant relation between institutional trading and spread. This implies that although both effective and percentage spreads have an inverse relation with institutional trading, inter stock effective spread is more dispersed than percentage spread. The opposite is true for the relation between spread and abnormal trading, and spread and abnormal trading volume. The correlation coefficient between effective spread and abnormal trading (trading volume) is $\hat{\rho} = .0297$ with p value .092 ($\hat{\rho} = -.0384$ with p value .029) is significant, but with respect to percentage spread both abnormal trading and abnormal trading volume show insignificant correlation. Note that in Table 2B, the mean correlation coefficient between

effective spread and abnormal trades (abnormal trading volume) is .0298 (-.0725) with 9 (12) firms reporting significant relations; for percentage spread, those means are -.0725 and -.0711 with 12 and 10 firms respectively reporting significant relations.

In summary, the highest numbers of significant relations are observed between spread and percentage spread (49), between abnormal trades and abnormal trading volume (46), and between volatility and abnormal trading (43). We interpret these three strongly positive correlations as follows. A strong correlation between spread and percentage spread is expected by design; the one between abnormal trading volume and abnormal trades confirm that certain traders who trade large volumes also trade frequently; and finally, volatility is more often related to frequency of trading than volume of trading (significant for 33 firms). The significantly positive relation between institutional trading and trading volume (significant for 15 firms) mildly support that institutions trade large volumes, however, there is an inverse relation between institutional trading and number of trades (significant for 14 firms) that casts some doubt on a direct relation between institutional trading, trading volume, and number of trades.

II.D. Regression Results of Panel Data Model

Table 3 reports the regression results for the panel data regressions with percentage spread¹³ as the dependent variable. In Model 1, all five independent variables namely abnormal number of trades (-0.006), abnormal trading volume (-0.013), price (-0.0289), volatility (0.152), and institutional proportion (-0.053), are significant at less

¹³ We have results from regression models using effective spread as the dependent variable. We do not report those since the results are virtually similar, except a few expected sign reversals for some coefficients, e.g., price.

than 1% level in determining percentage spread. The negative coefficients for abnormal trading and abnormal trading volume indicate support for trading cost minimization through increased activity arguments. We will expect an inverse relation between price and percentage spread since percentage spread is computed as effective spread over mid quote, a presumed linear function of price. This also suggests a direct relation between price inverse, an approximate measure of conditional beta risk (Miller and Scholes [1982]) and percentage spread. One would note though the remarkable similarities in the direction (signs) of the coefficients of abnormal trades, abnormal trading volume, and price particularly in the light of correlation reported in Table 2, Panel A, in which price does not have a significant correlation with either abnormal trading (number of trades) or abnormal trading volume and thus the information content of those variables maybe similar but independent of each other. The R^2 for the model is 88%.

In Model 2, we introduce two dummy variables for high and medium institutional proportions. The high (medium) institutional proportion variable has a value of one when the level of institutional trading is in the top (middle) 33% percentile, and zero otherwise. The coefficients for high dummy (-0.01), medium dummy (-0.006), and institutional proportion (-0.025) are all negative suggesting the negative slope becomes steeper (coefficient of high dummy is more negative than medium dummy) at a higher level of institutional trading. In Model 3, the institutional proportion variable is replaced by three variables – high, medium, and low proportion. The high (medium, low) proportion variable has the same value as institutional proportion if institutional trading proportion is in the top (middle, bottom) 33 percentile, and zero otherwise. The coefficients of all three variables are negative and significant at greater than 99 percent level of confidence

in the regression, but while the coefficient for low proportion is -0.069, that for the medium proportion is more negative at -0.073, but drops down substantially to -0.054 at the high proportion level. We interpret these results as follows. On average, there may be a mean positive effect on institutional trading embodied in the positive intercept. However, when there is an increase in institutional trading within a level, for percentage spread, the negative slope implying an inverse relation between percentage spread and institutional trading becomes steeper from low to medium level and then (coefficient of medium dummy is the most negative among all three levels) subsides at the highest level of institutional trading. This implies a convex, non-linear relation between institutional trading and percentage spread.

In model 4, we break up the institutional trading proportion into institutional buy and institutional sell to test for a difference in their effect on the spread. Results from this model show that spread reduces as institutional trading increases, be it buy or sell. The coefficients for buy and sell are significantly negative for both effective and percentage spread. We conduct an F-test (as per Neter et al [1996]) for the difference between the coefficients of buy and sell in this model recognizing that Model 4 is a full model and the corresponding reduced model is constructed by replacing institutional proportion with (institutional buy - institutional sell) in Model 1. Our computed F value for both effective and percentage spreads are less than 1.00 and thus we cannot reject the hypothesis that the coefficients of buy and sell in Model 4 are similar.

In order to provide some insight into the results so far we note the following. The coefficient estimates are consistent in all models for percentage spread i.e., the coefficients of abnormal trading (-), abnormal trading volume (-), price (-), and volatility

(+) have similar signs. Now if abnormal trading measured in number of trades and in trading volume denotes activity, then we would expect a negative coefficient for both abnormal number of trades, and abnormal trading volume. For percentage spread all those coefficients are negative as expected and consistent with trading cost literature and spread as a straddle option (Copeland and Galai [1983]). The signs for the coefficients of price (-) and volatility (+) as measures of risk are consistent with McInish and Wood (1992) and Heflin and Shaw (2000).

The consistently negative coefficients for institutional trading in all the models implying an inverse relation between institutional trading and spread is contrary to institutional trades revealing value related information. The inverse relation is contrary to the direct relation between spread and percent of block holding and institutional ownership reported by Heflin and Shaw (2000) and Sarin, Shastri, and Shastri (2002) respectively, but agree with Kini and Mian (1995) who report an inverse relation between institutional holding and spread. The inverse relation mirrors the relation between spread and abnormal trading volume and lends support to the effectiveness of using trading volume, and number of trades as proxies for institutional trades. However there are two caveats. First, the inverse relation between spread and institutional trading contradicts the direct relation between spread and trade size predicted by theories related to information content in trade size (Easley and O'Hara [1987]) and documented by Hasbrouck (1991) and LSB (1995). Second, if abnormal trading volume represents trading cost¹⁴ (an inverse relation with spread) then it is possible that institutional trading

¹⁴ The two leading arguments for an inverse relation between trading activity and spread are economies of scale in trading cost, and spread as a free straddle option provided by the market maker (Copeland and Galai (1983)).

affects spread in more than one way – trading cost, information, and possibly other (inventory) effects where the negative relations (spread decreasing) overpower the positive information effects of institutional trading. The trading cost effects may include economies of scale embedded in order processing cost, execution cost minimization strategy (Bertsimas and Lo [1998]), bid ask spread as a free straddle option (Copeland and Galai [1983]), and liquidity motive. We look into that possibility in the next three models, where we use an activity level dummy to differentiate between active vs. less active stocks and further in the decomposition of spread.

In models 5, 6, and 7, we add an indicator variable representing trading activity level of a firm. This may also represent firm size since larger firms are generally more active, have less private value related information because of their wide analysts coverage (Easley, O'Hara, and Paperman [1998]), and may have other non-information motives like liquidity. In model 5, the trading activity dummy is significant at greater than 99% confidence level. All other regressor variables in Model 5 are significant (at greater than 99% level) too, with model R^2 91%. A positive coefficient for the activity dummy variable in model 5 suggests that the intercept for actively traded firms is higher than less actively traded firms. Thus the mean percentage spread is higher than those of the less actively traded firms.

Model 6, with the interaction variable (trading activity dummy * institutional trading proportion) once again have all regressor variables significant at greater than 95% level of confidence. The interaction variable has negative coefficients in Model 6 (-.2302 with p-value <.0001). Notably though institutional trading has a positive coefficient (.0484 with a p-value <.0001) in model for percentage spread. To understand the

implications, let us look at the expected value of spreads for the actively traded and the less actively traded firms.

Holding everything else constant,

$$E(\%spread | \text{actively traded}) = (1.383 + .1595) - (.2302 - .0484) * inst\ trading$$

$$E(\%spread | \text{less actively traded}) = 1.383 + .0484 * inst\ trading .$$

Thus the mean percentage spread for actively traded firms is significantly greater than that of less actively traded firms. Further, percentage spread for less active firms increase as institutional trading increases, while for actively traded firms, it decreases as institutional trading increases. This finding is indicative of a non-linear relation between percentage spread and institutional trading where only the most actively traded (and perhaps expensive stocks) firms offer substantial inducements to institutions to trade heavily and benefit from low percentage spread. Bertsimas and Lo (1998) suggest lowering execution cost as a strategy for institutions. From information content in institutional trades perspective, the active, widely analysts' covered firms will have low information content (Easley, O' Hara, and Paperman [1998]) and thus market makers do not infer value related information about those securities from institutional trades; however, they infer information from institutional trading in less actively traded stocks for which information is rare and thus more costly.

Finally, in Model 7, we split up institutional trading into institutional buy and institutional sell. The model R^2 is 93% with all regressor variables significant at greater than 99% level of confidence. As in Model 4, the coefficients for institutional buy and sell are negative and significant implying that institutional trading *per se* and not the direction of trade reveals information to the market makers. For active stocks, we find

the following equation for expected percentage spread insightful in terms of decomposing the slope coefficient for institutional trading. Once again, holding everything else constant,

$$E(\%spread | \text{actively traded}) = (1.417 + .1292) - .1329(\text{inst trading}) - .0297(\text{inst.buy}) - .024(\text{inst.sell})$$

In the earlier paragraph, we found the coefficient for institutional trading for active stocks is approximately -.18 (-.2302 + .0484), which is broken down as -.13 due to institutional trading *per se*, -.03 due to institutional buy, and -.02 due to institutional sell respectively. This finding about irrelevance of direction of institutional trading is independent of whether the stocks are active or less active stocks emphasizes the role of gross order flow in the determination of spread.

In summary, the results strongly support our hypotheses (H1) that after controlling for activity, risk, and information through a variety of proxies institutional trading is a significant variable in the determination of both effective and percentage spreads. Thus institutional trading *per se* provides information (value related and non-value related) that is not revealed by abnormal number of trades, abnormal trading volume, total number of trades dummy, price, and volatility. The inverse relation between institutional trading and spread contradicts popular notion that institutional trades reveal value related information. We show that this inverse relation i.e. downward slope between percentage spread and institutional trading becomes steeper before it flattens out as proportion of institutional trading increases. We also find that the relation between spread and institutional trading is dependent on the trading activity level of stocks. Percentage spread increases for less actively traded stocks, and decreases for actively traded stocks for a similar increase in institutional trading proportion. Finally,

we can not reject the hypothesis that there is any difference in the relation between spread and institutional buy or sell and thus we claim that institutional trading *per se* drives the relation between spread and institutional trading, and the direction of institutional trades - buy or sell is redundant.

III. Analysis of the Components of the Spread

III.A. Institutional Trading and Components of Bid –Ask Spread

We extend our analysis of the relation between institutional trading and spread by decomposing the spread into its order processing and adverse selection components and investigating the effect of institutional trading on the individual components. We use a technique from LSB (1995) based on Stoll (1989) to decompose the spread into order processing and adverse selection components and hypothesize a relation between the individual components and institutional trading. Specifically, we test the prediction of Dey and Kazemi (2001) that there are two simultaneous pressures on the relation between institutional trading and spread – an increasing relation between the adverse selection component of institutional trading and spread and a decreasing relation between the liquidity component and spread.

We conjecture that if institutional trading combines information and liquidity effects then the information effect should increase the adverse selection component, the liquidity effect should decrease the order processing costs, and the net effect on spread may be positive or negative. We hypothesize the following functional forms for order processing and adverse selection costs component:

Order processing cost = f (trade size, gross order flow, institutional trading)

Adverse selection cost = f (trade size, net order flow, institutional trading).

In stipulating the above functional forms, we recognize that while gross volume (total order flow) is important for the determination of order processing cost (order processing cost is a typical cost function; cost is a function of volume), trade direction or net volume (buy volume - sell volume) is important in the determination of adverse selection costs. We therefore include log (buy/sell) as a variable in the regression model for the adverse selection component.

We determine through a set of simultaneous equations how institutional trading affects the order processing and the adverse selection components of the spread after controlling for number of trades, trading volume (order flow), and trade direction. Further we determine how the asymmetric information content and the liquidity motive in institutional buys and institutional sells affects the adverse selection (information) component of the spread. The simultaneous equations approach uses the cross correlation between the two regression equations to improve the estimates over OLS estimates. Stated as alternates, our hypotheses are:

H2a. The adverse-selection component should increase with institutional trading, and the order-processing component should decrease with institutional trading.

H2b. Institutional buys should have a differential effect on the adverse selection component of the spread from institutional sells.

III.B. Decomposition of Spread

In order to test the above hypotheses, we decompose the quoted spread for the firms in our sample to determine the adverse selection and order processing components. We follow Stoll (1989)¹⁵ decomposition technique used recently by LSB (1995) and Heflin and Shaw (2000). We choose this technique to ensure comparability of our results with those of LSB (1995) that documents a relation between the components of spread and trade size. The parameters $\hat{\lambda}$, and $\hat{\gamma}$ of the following regression models are estimates of the adverse selection and the order processing components of the spread.

$$\Delta Q_{t+1} = \lambda z_t + e_{t+1} \quad (2)$$

$$\Delta P_{t+1} = -\gamma z_t + u_{t+1} \quad (3)$$

where,

$P_t = \log(\text{trade price at time } t)$

$Q_t = \log(\text{quote mid point at time } t)$

$\Delta = \text{Change in the relative variable from } t \text{ to } t+1$

$z_t = P_t - Q_t$ and

$e_{t+1}, u_{t+1} = \text{Random error terms with zero mean and constant variance}$

$$E(e_{t+1}u_{t+1}) = 0.$$

We estimate λ and γ from OLS regression. Equations (2) and (3) are classical simultaneous equation to be estimated jointly; however, the assumption about zero

¹⁵There are several decomposition methodologies other than Stoll (1989) that are available in the literature. Huang and Stoll (1997), Glosten and Harris (1988), and George, Kaul and Nirmalendran (1991) provide estimates of the different components of the spread including adverse selection, order processing and inventory holding cost. As mentioned earlier, our empirical results are based on an information asymmetry model and are subject to the limitations that we ignore inventory holding cost component. Studies using different methodologies find the components of the spread to be different. Heflin and Shaw (2002) use both Stoll (1989) and Huang and Stoll (1997) decomposition and report no differential effect.

correlation between the error terms based on the reported findings in LSB (1995) make the OLS estimates similar to GLS estimates.

Table 4 reports descriptive statistics about the spread decomposition for the 65 firms in the sample¹⁶. We find that the order processing costs for our sample firms are higher, and adverse selection costs lower than that reported in LSB 1995. The means and medians of the components are similar indicating a symmetric (low skewness) distribution for the component costs. Not surprisingly, the descriptive statistics from our sample correspond closely with those from the largest volume sub-sample in LSB (1995) confirming the notion that institutions generally trade in large sizes in liquid stocks and also incur low execution cost.

III.C. SUR model

We use a cross sectional¹⁷ SUR (Seemingly Unrelated Regression) model to estimate the parameters of the following system of equations for the adverse selection and the order processing components of the daily spread for our sample firms.

¹⁶ Note that for this section, we have used the full sample of 65 firms and not 51 firms as in the panel regression. 14 of those 65 firms have less than 63 (61 or 62) days of data.

¹⁷ We use a cross sectional instead of a panel data model for the simultaneous regression. To ensure data adequacy for the regression models (2 and 3) to estimate the components of the spread, we ran the models for all 63 days together. All firms have on an average more than 20 trades a day, but most have less than 20 on a number of days within the sample period. Estimating the components of spread on a daily basis would mean using fewer than 20 data points for several days to run regression models 2 and 3.

$$OPC = \alpha_1 + \beta_{11}NumTrd + \beta_{12}InstitutionTrading + \beta_{13}AvgVol + e_o \quad (4)$$

$$ADV = \alpha_2 + \beta_{21}NumTrd + \beta_{22}InstitutionTrading + \beta_{23}AvNetVol + e_a \quad (5)$$

where,

OPC = Order Processing cost

ADV = Adverse Selection cost

NumTrd = Mean daily number of trades for firm

InstitutionTrading = Mean institutional trading proportion for firm

AvgVol = Log(Average daily trade volume for firm)

AvNetVol = Log(Average buy volume per trade/Average sell volume per trade).

The SUR model assumes that

$$E(e_o) = E(e_a) = 0.$$

$$E(e_{oi}e_{aj}) = \sigma_{oa} \text{ if } i = j \text{ and } 0 \text{ otherwise for } i, j \in (1..65).$$

We also test the hypothesis that institutional buys and sells have a differential impact on adverse selection and order processing costs by breaking down institutional trading in equations 4 and 5 into institutional buys and sells. We compute an F statistic to test the null hypothesis that the coefficients of institutional buy and institutional sell are similar.

III.D. Results of the SUR model

Table 5 reports the results of the SUR model. The number of observations for each system of equations estimated is 65 – one observation for each stock in the sample. The system wide R² for the two systems of equations reported in Table 5 are .301 and .302 respectively. Our results show that institutional trading significantly increases the

adverse selection component of the spread but reduces the order processing cost component. This suggests that the market maker acknowledges institutional trading to have multiple trading motives i.e., information, liquidity, and trading cost minimization and factors in all of those in the determination of the spread. The order flow variable (AvNetVol) is significantly negative, thus an increase in buys has the effect of reducing the spread. The “number of trades” variable is significantly negative in the adverse selection equation and positive in the order processing equation – indicating its dichotomous effect on the spread components. The volume variable is negative in the order processing equation as expected. As volume increases, order processing costs decrease. We interpret this result as supporting the assumption of a classical decreasing marginal cost function for order processing costs due to economies of scale.

For our second model, we estimate the system of equations with institutional buys and sells in place of total institutional proportion. We find that institutional sells are positive and significant (at greater than 90% level of confidence), but institutional buys are insignificant in determining the adverse component. We test if the coefficients of buy and sell are similar. Our computed F value is 11.5 and thus we reject the hypothesis that the coefficients of institutional buy and institutional sell are similar at greater than 99% level of confidence. One explanation for this arises from the asymmetry between buy and sell recommendations from analysts. Barber et al (2001) report 29.8%, 24.3%, 39.5%, 3.9%, and 2.6% of recommendations from security analysts are strong buy, buy, hold, sell, and strong sell recommendations respectively. Institutional buys do not convey any new information since they seem to match publicly available analysts’ recommendations that are predominantly buys. However, since recommendations are so

rare, institutional sells provide relevant information to the market maker who builds it into the price discovery process and the bid ask spread.

IV. Conclusion

The role of institutional trading in the determination of spread is of interest to regulators and market makers. However there is no existing empirical evidence on the relation between institutional trading and spread. Using transactions data, we present empirical evidence that institutional trading *per se* affects spread i.e., institutional trading has power to explain the variability of spread beyond those contained in trading frequency, trading volume, price and volatility. We find the net effect of increased institutional trading to be a reduction in spread. However, the relation between spread and institutional trading is not linear and depends on the level of institutional trading and whether the stocks are active or less active stocks. Percentage spread decreases at different rates at different levels of institutional trading, with the steepest decline being at mid level, and flattening out at the highest level of institutional trading. Further analysis reveals that percentage spread for less active stocks is initially lower than that of active stocks but increases, while for active stocks it decreases as institutional trading increases. Our analysis indicates that institutional buys and sells affect spread similarly; however, institutional buys seem to reduce spreads, but sells seem to increase the adverse selection component.

Bibliography

- Barber, Brad, Reuben Lehavy, Maureen McNichols, and Brett Trueman, 2001, Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns, *Journal of Finance*, 54, 531-63
- Barber, Brad, and Terance Odean, 2000, Trading is Hazardous to your Wealth: The Common Stock Investment Performance of Individual Investors, *Journal of Finance*, 55, 773-806
- Branch, Ben, and Walter Freed, 1977, Bid-Asked Spreads on the AMEX and the Big Board, *Journal of Finance*, v32, n1, 159-163
- Bertsimas, Dimitris, and Andrew Lo, 1998, Optimal Control of Execution Costs, *Journal of Financial Markets*, v1, n1, 1-50
- Chakrovarty, Sugato, 2001, Stealth Trading: Which Traders' Move Stock Prices, *Journal of Financial Economics*, 61(2), 289-307
- Chung, Kee, and C. Charoenwong, 1998, Insider Trading and the Bid-Ask Spread, *Financial Review*, v33, n3, 1-20
- Copeland, Tom, and Dan Galai, 1983, Information Effects on the Bid-Ask Spread, *Journal of Finance*, 38, 1457-69
- Demsetz, H., 1968, The Cost of Transacting, *Quarterly Journal of Economics*, v82, 33-53
- Dey, Malay K. and H. Kazemi, 2001, Bid-ask Spread in a Competitive Market with Block Traders and Order Size, University of Massachusetts, *Working Paper*
- Easley, D, Maureen O'Hara, 1987, Price, Trade Size, And Information in Securities Markets, *Journal of Financial Economics*, v19, 69-90
- Easley, D, Maureen O'Hara, and J. Paperman, 1998, Financial Analysis and Information Based Trade, *Journal of Financial Markets*, 1(2), 175-202
- George Thomas, G. Kaul, and Nirmalendran, 1991, Estimation of the Bid-Ask Spread and its Components: A New Approach, *Review of Financial Studies*, v4, n4, 623-56
- Glosten, Larry, and Lawrence Harris, 1988, Estimating the Components of Bid /Ask Spread, *Journal of Financial Economics*, v21, n1, 123-42
- Harris, Lawrence, 1994, Minimum Price Variations, Discrete Bid/Ask Spreads and Quotations Sizes, *Review of Financial Studies*, 7(1), 149-78

- Hasbrouck, Joel, 1991, Measuring the Information Content of Stock Trades, *Journal of Finance*, v46, n1, 179-207
- Hasbrouck, Joel, 1992, Using the TORQ Database, *NYSE Working Paper*
- Heflin, Frank, and Kenneth Shaw, 2000, Blockholder Ownership and Market Liquidity, *Journal of Financial and Quantitative Analysis*, v35, n4, 621-33
- Huang, Roger and Hans Stoll, 1997, The Components of the Bid-Ask spread: A General Approach, *Review of Financial Studies*, v10, n4, 995-1034
- Kennedy, P., 1998, *A Guide to Econometrics*, 4th edition, The MIT Press, MA
- Kim, Sung-Hun, and Joseph Ogden, 1996, Determinants of the Components of Bid-Ask Spreads on Stock, *European Financial Management*, v1, 127-45
- Kini, Omesh, and S. Mian, 1995, Bid-Ask spread and Ownership Structure, *Journal of Financial Research*, v18, 401-14
- Klock, Mark, and Timothy McCormick, 1999, The Impact of Market Maker Competition on Nasdaq Spreads, *The Financial Review*, v34, 55-74
- Koski, Jennifer L. and R. Michaely, 2000, Prices, Liquidity, and the Information Content of Trades, *Review of Financial Studies*, v13, n3, 659-96
- Koski, Jennifer, and John Scruggs, 1998, Who Trades Around the Ex-dividend Day? Evidence From NYSE Audit File Data, *Financial Management*, 27, 58-72
- Laux, Paul, 1993, Trade Sizes and Theories of the Bid-Ask Spread, *Journal of Financial Research*, v16, n3, 237-49.
- Laux, Paul, 1995, Dealer Market structure, Outside competition, and Bid Ask Spread, *Journal of Economic Dynamics and Control*, v19, 683-710.
- Lee, Charles, and B. Radhakrishna, 2000, Inferring Investor Behavior: Evidence from TORQ Data, *Journal of Financial Markets*, v3, 2, 183-204
- Lee, Charles, and M. Ready, 1991, Inferring Trade Directions from Intraday Data, *Journal of Finance*, v46, n3, 733-46
- Lin, Ji-chai, Gary Sanger, and Geoffrey Booth, 1995, Trade Size and Components of the Bid-Ask Spread, *Review of Financial Studies*, v8, n4, 1153-83
- McInish Thomas, and Robert Wood, 1992, An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE stocks, *The Journal of Finance*, v47, n2, 753-64

- Miller, Merton, and Myron Scholes, 1982, Dividend and Taxes: Some Empirical Evidence, *Journal of Political Economy*, 90, 1118-41
- Neter, J, M. Kutner, C. Nachtsheim, and W. Wasserman, 1996, *Applied Linear Regression Models*, 3rd Edition, Irwin-McGraw Hill.
- Parks, R.W., 1967, Efficient Estimation of a System of Regression Equations when Disturbances are Both Serially and Contemporaneously Correlated, *Journal of American Statistical Association*, v62, 500-509
- Peles, Nadav, 1992, Contemporaneous Correlations Within Equity Trading of Institutional Investors, *Working Paper*, Columbia University
- Radhakrishna, B., 1995, Investor Heterogeneity and Earnings Announcement, *Ph.D. Dissertation*, University of Michigan, Ann Arbor
- Sarin, A, Karen Shastri, and Kuldeep Shastri, 2000, Ownership Structure and Stock Market Liquidity, *SSRN Working Paper*
- Schwartz, Robert, 1988, *Equity Markets*, Harper and Row, New York
- Scholes, Myron, 1972, The Market for Securities: Substitution Versus Price Pressure and the Effects of Information on Share Prices, *Journal of Business*, 45, 179-211
- Sias, R., L. Starks, and S. Titman, 2001, The Price Impact of Institutional Trading, *Working Paper*
- Stoll, Hans, 1989, Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests, *Journal of Finance*, v44, n1, 115-34

Table 1**Descriptive Statistics of Variables Used in Panel Regressions**

	Mean	Median	Quartile Range	Minimum	Maximum	Standard Deviation ⁷
Effective Spread ¹	0.126	0.120	0.020	0.019	0.514	0.023
Percentage Spread ¹	0.773	0.613	0.167	0.115	6.022	0.167
Institutional Trading Proportion ²	0.299	0.296	0.060	0.000	1.000	0.100
Number of Trades ³	1.000	0.942	0.188	0.095	3.924	0.415
Trade Volume ⁴	1.000	0.793	0.177	0.027	12.351	0.821
Price ⁵	24.716	22.000	14.000	2.500	83.500	1.200
Volatility ⁶	0.030	0.0125	0.0232	0.000	1.727	0.039

¹Effective spread is defined as $\text{Abs}[(\text{Price} - \text{Mid Quote}) * 2]$. The variable used in the panel regressions is average spread per day for each stock. Percentage Spread is computed as effective spread/mid quote * 100.

²Institutional trading proportion is computed as trades by institutions as a proportion of total trades for a day.

³Number of trades is computed as the number of trades for each stock each day divided by average trades per day over the entire trading period. Thus, the mean of this variable on a firm-by-firm basis is by design equal to 1.

⁴Trade Volume is computed as share volume of each stock for each day divided by average share volume per day over the entire trading period. Thus, the mean of this variable on a firm-by-firm basis is by design equal to 1. The standard deviation presented here is the mean of the within stock standard deviations.

⁵Price is closing price of a stock for the day.

⁶Volatility is the realized intra-day standard deviation of prices.

⁷The standard deviation presented here is the mean of the within stock standard deviations.

Table 2
Correlation Analysis

PANEL A: Pooled sample Pearson Correlations (N=3213) ^a

Variables	Spread	% Spread	NTradePct	InstProp	DlyAvg	Price ⁵	Volatility
Spread ¹	1.000						
Percentage Spread	.0905 ($<.0001$)	1.000					
NTradePct ²	.0297 (.092)	-.0163 (.355)	1.000				
InstProp ³	-.0005 (.979)	-.2617 ($<.0001$)	-.0880 ($<.0001$)	1.000			
DlyAvg ⁴	-.0384 (.029)	-.0184 (.297)	.5164 ($<.0001$)	.0994 ($<.0001$)	1.000		
Price ⁵	.1319 ($<.0001$)	-.6296 ($<.0001$)	-.0003 (.854)	.3032 ($<.0001$)	-.0025 (.8858)	1.000	
Volatility ⁶	.1884 ($<.0001$)	-.1481 ($<.0001$)	0.2539 ($<.0001$)	.1369 ($<.0001$)	.1615 ($<.0001$)	.3125 ($<.0001$)	1.000

^aPanel A table values are 1) Parameter estimates, and 2) P-Values in parentheses. The variables in Panels A and B in Table 2 are defined as follows.

Variable Descriptions:

¹**Spread** = Effective spread is defined as $\text{Abs}[(\text{Price} - \text{Mid Quote}) * 2]$. The variable used in the panel regressions is average spread per day for each firm. Percentage Spread is computed as effective spread/mid quote *100.

²**NTradePct** = Number of trades - This is computed as the number of trades for each stock each day divided by average trades per day over the entire trading period.

³**InstProp** = Institutional trading proportion – This is computed as trades by institutions as a proportion of total trades for a day.

⁴**DlyAvg** = Daily Trade Volume - This is computed as trading volume of each stock for each day divided by average trading volume per day over the entire trading period.

⁵**Price** = Closing Price.

⁶**Volatility** = Volatility is the realized intra-day standard deviation of prices.

PANEL B: Mean of firm Correlations^a

Variables	Spread	% Spread	NtradePct	InstProp	DlyAvg	Price ⁵	Volatility
Spread ¹	1.000						
Percentage Spread	.8377 (49)	1.000					
NTradePct ²	.0298 (9)	-.0186 (16)	1.000				
InstProp ³	-.0954 (13)	-.0593 (14)	-.1047 (14)	1.000			
DlyAvg ⁴	-.0725 (12)	-.0711 (10)	.5605 (46)	.1231 (15)	1.000		
Price ⁵	.0164 (13)	-.2725 (30)	.0244 (23)	-.0186 (10)	-.0163 (19)	1.000	
Volatility ⁶	.1106 (16)	.0596 (16)	.3672 (43)	.0432 (6)	.2944 (33)	.0366 (20)	1.000

Panel B table values are 1) Parameter estimates based on mean of firm estimates, 2) the number of firms for which coefficients are significant at p-values of less than 10% inside parentheses. The mean coefficient is computed from the 51 firms in the sample.

Table 3
Panel Regression Results

Dependent Variable: Percentage Spread (for all models)

Model	1	2	3	4	5	6	7
Intercept	1.454 ($<.0001$)	1.4502 ($<.0001$)	1.4548 ($<.0001$)	1.4621 ($<.0001$)	1.417 (.0001)	1.383 ($<.0001$)	1.417 ($<.001$)
Nos.trade Pct	-.006 (.0022)	-.0053 (.0075)	-.0062 (.0016)	-.0075 ($<.0001$)	-.0067 (.0004)	-.0074 ($<.0001$)	-.0104 ($<.0001$)
Dly_Avg Volume	-.013 ($<.0001$)	-.013 ($<.0001$)	-.013 ($<.0001$)	-.0125 ($<.0001$)	-.0124 ($<.0001$)	-.013 ($<.0001$)	-.0113 ($<.0001$)
Price	-.0289 ($<.0001$)	-.0288 ($<.0001$)	-.0287 ($<.0001$)	-.0282 ($<.0001$)	-.0282 ($<.0001$)	-.028 ($<.0001$)	-.028 ($<.0001$)
Volatility	.152 ($<.0001$)	.158 ($<.0001$)	.1552 ($<.0001$)	.1517 ($<.0001$)	.1484 ($<.0001$)	.1596 ($<.0001$)	.1538 ($<.0001$)
InstProp	-.053 ($<.0001$)	-.0252 (.0037)			-.0539 ($<.0001$)	.0484 ($<.0001$)	
InstBuy				-.0521 ($<.0001$)			-.0297 ($<.0001$)
InstSell				-.0494 ($<.0001$)			-.024 ($<.0001$)
High- Dum		-.01 ($<.0001$)					
Med- Dum		-.0058 (.0021)					
High InstProp			-.0541 ($<.0001$)				
Med Instprop			-.0727 ($<.0001$)				
Low Instprop			-.0687 ($<.0001$)				
Tradedum					.0911 ($<.0001$)	.1595 ($<.0001$)	.1292 ($<.0001$)
(Tradedum *Instprop)						-.2302 ($<.0001$)	-.1329 ($<.0001$)
R ²	.882	.886	.888	.893	.907	.898	.932

All regression models in Tables 3A and 3B above are run with a panel data set that has 51 firms each with 63 trading days. The basic model for percentage spread given in Model 1 is,

$$Spr_{nt} = \alpha + \beta_1 NtrPct_{nt} + \beta_2 DlyAv_{nt} + \beta_3 Pr_{nt} + \beta_4 Volty_{nt} + \beta_5 Instpr_{nt} + \varepsilon_{nt}$$
All other models are extensions of the basic model. The Parks (1967) method with $AR(1)$ and contemporaneously correlated error term is used for estimation. Table values are parameter estimates with p-values in parentheses.

Variable Descriptions:

Spread = Effective spread is defined as $Abs[(Price - Mid Quote)*2]$. Percentage Spread is computed as effective spread/mid quote *100. The variable used in the panel regressions is average spread per day for each firm.

NTradePct = Number of trades - This is computed as the number of trades for each stock each day divided by average trades per day over the entire trading period.

DlyAvg Volume = Daily Trade Volume - This is computed as share volume of each stock for each day divided by average share volume per day over the entire trading period.

Price = Closing Price. Using average price and price inverse do not change the regression estimates qualitatively.

Volatility = Intra day realized standard deviation of price. Using return volatility does not change the regression estimates qualitatively.

InstProp = Institutional trading proportion – This is computed as trades by institutions as a proportion of total trades for a day. **HighInstProp** has the value of InstProp if the institutional proportion is in the top third percentile of institutional proportions, and 0 otherwise. **MedInstprop** has the value of InstProp if the institutional proportion is in the middle third percentile of institutional proportions, and 0 otherwise. **LowInstprop** has the value of InstProp if the institutional proportion is in the bottom third percentile of institutional proportions, and 0 otherwise.

HighDum has the value 1 if the institutional proportion is in the top third percentile of institutional proportions, and 0 otherwise. **MedDum** has the value 1 if the institutional proportion is in the middle third percentile of institutional proportions, and 0 otherwise.

InstBuy = Institutional Buy trading proportion – This is computed as #buys by institutions as a proportion of total trades for a day.

InstSell = Institutional Sell trading proportion – This is computed as #sells by institutions as a proportion of total trades for a day.

Tradedum has the value 1, if average daily trade for a stock is greater than 45, the median number of daily trades; 0 otherwise.

Tradedum*Instprop = Tradedum, the dummy variable with value 1 or 0 times institutional proportion.

Table 4

**Descriptive Statistics: Components of the spread
as percentages of spread**

	Mean	Median	Minimum	Maximum	Standard Deviation
Adverse Selection Component ¹	25.3%	25.2%	2.8 %	46.1%	11.5%
Order Processing Component	40.1%	39.2%	13.4%	79.8%	15.9%

¹The effective spread is decomposed into its components using methodology originally proposed by Stoll (1989) and used recently by Lin, Sanger and Booth (1995). The descriptive statistics presented above is for the cross-section of 65 firms in the sample and is calculated from the mean proportion for each stock in our sample, of the adverse selection component and order processing component of the spread.

Table 5
Regression Results of System of Equations with Adverse Selection and Order Processing Costs on Trades, Volume and Institutional Trading

Dependent Variable	Intercept	NumTrd	Institution Trading	Institution Buy	Institution Sell	AvNetVol	AvgVol
OPC	.1817 (<.0001)	.00007 (.005)	-.0507 (.0798)				-.0107 (.0008)
ADV	.0137 (.0671)	-.00005 (.0007)	.0740 (.0022)			-.0131 (.0172)	
OPC	.1804 (<.0001)	.00007 (.005)	-.0511 (.0776)				-.0106 (.001)
ADV	.0125 (.1079)	-.00005 (.0008)		.0494 (.2969)	.1079 (.0814)	-.0135 (.0157)	

All regressions are run as seemingly unrelated regressions (SUR). The data set has observations for 65 firms for which data is available. Table values are parameter estimates with p-values in parentheses. **The system wide R² are .301 and .302 respectively for the two sets of equations above.**

Variable Descriptions: (All averages are computed over the entire 3-month trading period in the TORQ data set on a firm-by-firm basis.)

OPV = Order Processing component of the average effective spread for firm.

ADV = Adverse Selection component of the average effective spread for firm.

NumTrd = Average trades per day.

Institutional Trading = Average institutional proportion of trades.

Institution Buy = Average institutional buy proportion of trades.

institution Sell = Average institutional sell proportion of trades.

AvNetVol = Natural Log of [Average Buy Volume per trade/Average Sell Volume per trade]

AvgVol = Natural Log of Average Daily Trade Volume.