Caught On Tape: Predicting Institutional Ownership With Order Flow*

John Y. Campbell, Tarun Ramadorai and Tuomo O. Vuolteenaho†

March 2004

Abstract

Many questions about institutional trading behavior can only be answered if one can track institutional equity ownership continuously, yet institutional ownership data are only available on quarterly reporting dates. We infer institutional trading behavior from the "tape", the Transactions and Quotes database of the New York Stock Exchange, by regressing quarterly changes in reported institutional ownership on quarterly buy and sell volume in different trade size categories. We find that institutions in aggregate demand liquidity, in that total buy (sell) volume predicts increasing (decreasing) institutional ownership. Institutions also tend to trade in large or very small sizes, in that buy (sell) volume at these sizes predicts increasing (decreasing) institutional ownership, while the pattern reverses at intermediate trade sizes that are favored by individuals. Our regression method predicts institutional ownership significantly better than the simple cutoff rules used in previous research.

*We thank Peter Hawthorne, Sung Seo, and especially Allie Schwartz for excellent research assistance, and Boris Kovtunenko and Nathan Sosner for their assistance with the Spectrum dataset. Thanks also to Pablo Casas-Arce and Jos van Bommel for useful discussions. This material is based upon work supported by the National Science Foundation under Grant No. 0214061 to Campbell, and by Morgan Stanley and Co. under its Microstructure Research Program.

†John Campbell and Tuomo Vuolteenaho are at Harvard University, Department of Economics, Littauer Center, Cambridge, MA 02138, USA, john_campbell@harvard.edu and t_vuolteenaho@harvard.edu, and the NBER. Tarun Ramadorai is at the University of Oxford, Said Business School, Park End St., Oxford OX1 1HP, UK, tarun.ramadorai@said-business-school.oxford.ac.uk.

1. Introduction

How do institutional investors trade in equity markets? Do they hold stocks that deliver high average returns? Do they arbitrage irrationalities in individual investors' responses to information? Are they a stabilizing or destabilizing influence on stock prices? These questions have been the focus of a large and recent body of empirical literature.

Lakonishok, Shleifer, and Vishny (1992), Grinblatt, Titman, and Wermers (1995), Wermers (1999, 2000), Nofsinger and Sias (1999), and Grinblatt and Keloharju (2000a, b) show that quarterly increases in institutional ownership and quarterly stock returns are contemporaneously correlated. Several studies investigate this relationship further, and find evidence that short-term expected returns are higher (lower) for stocks that have recently been subject to significant institutional buying (selling).[1] Some authors, notably Lakonishok, Shleifer, and Vishny (1992), suggest that institutional investors follow simple price-momentum strategies that push stock prices away from fundamental values. This is disputed by others, such as Cohen, Gompers, and Vuolteenaho (2002), who find that institutions are not simply following price-momentum strategies; rather, they sell shares to individuals when a stock price increases in the absence of any news about underlying cash flows.

One limitation of this literature is that it is difficult to measure changes in institutional ownership as they occur. While some countries, such as Finland, do record institutional ownership continuously, in the United States institutional positions are reported only quarterly in 13-F filings to the Securities and Exchange Commission. A quarterly data frequency makes it hard to say whether institutions are reacting to stock price movements or causing price movements, and makes it impossible to measure institutional responses to high-frequency news such as earnings announcements.

To measure institutional trading at high frequency, some authors have looked at data on equity transactions, available on the New York Stock Exchange Trade and Quotations (TAQ) database. Most transactions can be identified as buy orders or sell orders using the

[1] See Daniel, Grinblatt, Titman, and Wermers (1997), Wermers (1999), Chen, Jegadeesh, and Wermers (2000), and Gompers and Metrick (2001), among others.

procedure of Lee and Ready (1991), which compares the transaction price to posted bid and ask quotes.

A more difficult challenge is to identify orders as coming from institutions or individuals. A common procedure is to label orders above some upper cutoff size as institutional, and those below a lower cutoff size as individual. Trades at intermediate sizes remain unclassified. Lee and Radhakrishna (2000) evaluate several alternative cutoff rules by applying them to the TORQ dataset, a sample of trades with complete identification of market participants. They find, for example, that upper and lower cutoffs of \$20,000 and \$2,500 are most effective at accurately classifying trades in small stocks. Unfortunately the TORQ dataset includes only 144 stocks over a three-month period in 1994 and it is not clear that these results apply more generally or in more recent data.

In this paper we develop a new method for inferring high-frequency institutional trading behavior. Our method combines two datasets that in the past have been used separately in analyses of investor behavior. The TAQ database gives us trade-by-trade data pertaining to all listed stocks on the NYSE and AMEX, NASDAQ national market system, and small cap stocks, beginning in 1993. We restrict the current analysis to stocks traded on the NYSE and AMEX. TAQ is essentially the "tape", recording transactions prices and quantities of every trade conducted on these exchanges. We match TAQ to the Spectrum database. Spectrum records the SEC mandated 13-F filings of large institutional investors, providing quarterly snapshots of institutional holdings. We use the cumulative quarterly trades recorded on the "tape" to predict institutional holdings in Spectrum. By regressing changes in institutional ownership on cumulative trades of different sizes, we find the best function mapping trade size to institutional behavior. This function can be used to track institutional trading on a daily or intra-daily basis.

Our first finding is that institutions on average demand liquidity. Across all trades (ignoring trade sizes), volume classifiable as buys predicts an increase and volume classifiable as sells predicts a decline in reported institutional ownership. Thus, we conclude that institutions use the liquidity provided by the specialist and possibly also provided by limit orders from individuals.

Second, we find that buying at the ask and selling at the bid is more likely to be due to institutions if the trade size is either very small or very large. Trades that are either under \$2,000 or over \$30,000 in size are very likely to be initiated by institutions, whereas intermediate size trades are relatively more likely to be by individuals.

Our third finding is that our method of inferring institutional buying and selling from the "tape" significantly outperforms the simple classification rules in previous literature. For example, a simple cut-off rule that classifies all trades over \$20,000 as institutional has a negative R^2 when used as a predictor of the change in institutional ownership. This is in contrast to the 10 percent R^2 obtained by our method.

The organization of the paper is as follows. Section 2 introduces the TAQ, Spectrum and CRSP data used in the study, and conducts a preliminary data analysis. Section 3 presents and applies our method for predicting institutional ownership. Section 4 concludes.

2. Preliminary data analysis

2.1. CRSP data

Shares outstanding, stock returns, share codes, exchange codes and prices for all stocks come from the Center for Research on Security Prices (CRSP) daily and monthly files. In the current analysis, we focus on ordinary common shares of firms incorporated in the United States that traded on the NYSE and AMEX. Our sample begins in January 1993, and ends in December 2000. We use the CRSP PERMNO, a permanent number assigned to each security, to match CRSP data to TAQ and Spectrum data. Figure 1 shows the evolution of the number of matched firms in our data over time. The maximum number of firms is 2222, in the third quarter of 1998. The minimum number of firms is 1843, in the first quarter of 1993.

In the majority of our analysis, we present results for all firms, as well as for five quintiles of firms, where quintile breakpoints and membership are determined by the market capitalization (size) of a firm at the start of each quarter. Our data are filtered carefully, as described below. After filtering, our final sample consists of 3402 firms. When sorted quarterly into size quintiles, this results in 744 firms in the largest quintile, and between 1194 and 1422 firms in the other four quintiles (these numbers include transitions of firms between quintiles), and 66,805 firm quarters in total.

2.2. TAQ data

The Transactions and Quotes (TAQ) Database of the New York Stock Exchange contains trade-by-trade data pertaining to all listed stocks, beginning in 1993. TAQ records transactions prices and quantities of all trades, as well as a record of all stock price quotes that were made. TAQ lists stocks by their tickers. We map each ticker symbol to a CRSP PERMNO. As tickers change over time, and are sometimes recycled or reassigned, this mapping changes over time.

The TAQ database does not classify transactions as buys or sells. To classify the direction of trade, we use an algorithm suggested by Lee and Ready (1991). This algorithm looks at the price of each stock trade relative to contemporaneous quotes in the same stock to determine whether a transaction is a buy or sell. In cases where this trade-quote comparison cannot be accomplished, the algorithm classifies trades that take place on an uptick as buys, and trades that take place on a downtick as sells. The Lee-Ready algorithm cannot classify some trades, including those executed at the opening auction of the NYSE, trades which are labelled as having been batched or split up in execution, and cancelled trades. We aggregate all these trades, together with "zero-tick" trades which cannot be reliably identified as buys or sells, into a separate bin, and use this bin of unclassifiable trades as an additional input into our prediction exercise.

Lee and Radhakrishna (2000) find that the Lee-Ready classification of buys and sells is highly accurate; however it will inevitably misclassify some trades which will create measurement error in our data. Appendix 1 describes in greater detail our implementation of the Lee-Ready algorithm.

Once we have classified trades as buys or sells, we assign them to bins based on their dollar size. In all, we have 19 size bins whose lower cutoffs are \$0, \$2000, \$3000, \$5000, \$7000, \$9000, \$10,000, \$20,000, \$30,000, \$50,000, \$70,000, \$90,000, \$100,000, \$200,000, \$300,000, \$500,000, \$700,000, \$900,000, and \$1 million. In several of our specifications below, we use buy and sell bins separately, and in others, we subtract sells from buys to get the net order flow within each trade size bin. We aggregate all shares traded in these dollar size bins to the daily frequency, and then normalize each daily bin by the daily shares outstanding as reported in the CRSP database. This procedure ensures that our results are not distorted by stock splits.

We aggregate the daily normalized trades within each quarter to obtain quarterly buy and sell volume at each trade size. The difference between these is net order imbalance or net order flow. We normalize and aggregate unclassifiable volume in a similar fashion. The sum of buy, sell, and unclassifiable volumes is the TAQ measure of total volume in each stock-quarter.

We filter the data in order to eliminate potential sources of error. We first exclude all stock-quarters for which TAQ total volume as a percentage of shares outstanding is greater than 200 percent (there are very few such observations). We then compute the standard deviation across stock-quarters of each volume measure and the net order imbalance, relative to each quarter's cross-sectional mean, and winsorize all observations that are further than 2.5 standard deviations from their cross-sectional mean. That is, we replace such outliers with the cross-sectional mean for the quarter plus or minus 2.5 standard deviations. This winsorization procedure affects between 2.50 and 3.15 percent of our data.

Figure 2 shows the cross-sectional mean and standard deviation of TAQ total volume as a percentage of shares outstanding in each quarter, in annualized percentage points. In the early years of our sample period total volume averaged between 60 percent and 80 percent of shares outstanding per year; this increased to between 80 percent and 100 percent in the later years of the sample. These numbers are consistent with other recent studies such as Chen, Hong and Stein (2002) and Daves, Wansley and Zhang (2003).[2] There is considerable cross-sectional heterogeneity in volume as illustrated by the cross-sectional standard deviation of 30 percent to 40 percent.

Some of this cross-sectional heterogeneity can be explained by differences in the trading patterns in small and large stocks. Table I reports means, medians, and standard deviations across all firm-quarters, and across firm-quarters within each quintile of market capitalization. Mean total volume ranges from 53 percent of shares outstanding in the smallest quintile to 91 percent in the largest quintile. The distribution of total volume is positively skewed within each quintile, so median volumes are somewhat lower but also increase with market capitalization. The within-quintile annualized standard deviations (computed by multiplying quarterly standard deviation by a factor of 200, under the assumption that quarterly observations are iid) are fairly similar for stocks of all sizes, ranging from 27 percent to 33 percent.

Table I also reports the moments of the net order flow for each size quintile. Mean net order flow increases strongly with market capitalization, ranging from –2.1 percent for the smallest quintile to 4.5 percent for the largest quintile. This suggests that over our sample period, there has been buying pressure in large stocks and selling pressure in small stocks, with the opposite side of the transactions being accommodated by unclassifiable trades that might include limit orders.[3] This is consistent with the strong price performance of large stocks during most of this period.

Unclassifiable volume is on average about 15 percent of shares outstanding in our dataset. This number increases with firm size roughly in proportion to total volume; our algorithm fails to classify 18 percent of total volume in the smallest quintile, and 21 percent of total volume in the largest quintile. It is encouraging that the algorithm appears equally reliable

[2] These cross-sectional moments are calculated on an equal-weighted basis across firms, so to the extent that volume varies systematically with firm size, our measures will not correspond to total volume as calculated for the exchange as a whole. Our measures represent the behaviour of the average firm.

[3] In support of this interpretation, net order flow is strongly negatively correlated with Greene's [1995] measure of limit order depth for all size quintiles of stocks. This measure essentially identifies a limit order execution as the quoted depth when a market order execution is accompanied by a movement of the revised quote away from the quoted midpoint.

among firms of all sizes. Note that the means of buy volume, sell volume, and unclassifiable volume do not exactly sum to the mean of total volume because each of these variables has been winsorized separately.

Figure 2 summarizes the distribution of buy and sell volume across trade sizes. The figure reports three histograms for the smallest quintile, the median quintile, and the largest quintile of stocks. Since our trade size bins have different widths, ranging from $1000 in the second bin to $200,000 in the penultimate bin and even more in the largest bin, we normalize each percentage of total buy or sell volume by the width of each bin, plotting "trade intensities" rather than trade sizes within each bin. As the largest bin aggregates all trades greater than $1 million in size, we arbitrarily assume that this bin has a width of $5 million.

It is immediately obvious from Figure 2 that trade sizes are positively skewed, and that their distribution varies strongly with the market capitalization of the firm. In the smallest quintile of stocks almost no trades of over $70,000 are observed, while such large trades are commonplace in the largest quintile of stocks. A more subtle pattern is that in small stocks, buys tend to be somewhat smaller than sells, while in large stocks the reverse is true.

Table II summarizes the distribution of trade sizes in a somewhat different fashion. The table reports the medians and cross-sectional standard deviations of total classifiable volume (buys plus sells) in each trade size bin for each quintile of market capitalization. The rarity of large trades in small stocks is apparent in the zero medians and tiny standard deviations for large-size volume in the smallest quintile of firms.

2.3. Spectrum data

Our data on institutional equity ownership come from the Spectrum database, currently distributed by Thomson Financial. They have been extensively cleaned by Kovtunenko and Sosner (2003) to remove inconsistencies, and to fill in missing information that can be reconstructed from prior and future Spectrum observations for the same stock. A more detailed description of the Spectrum data is presented in Appendix 2. Again, we first filter the data by removing any observation for which the change in Spectrum recorded institutional ownership as a percentage of firm shares outstanding is greater than 100 percent (there are very few such observations). We then winsorize these data in the same manner as the TAQ data, truncating observations that are more than 2.5 standard deviations away from each quarter's cross-sectional mean.

Table I reports the mean, median, and standard deviation of the change in institutional ownership, as a percentage of shares outstanding. Across all firms, institutional ownership increased by an average of 0.6 percent per year, but this overall trend conceals a shift by institutions from small firms to large and especially mid-cap firms. Institutional ownership fell by 1.3 percent per year in the smallest quintile but rose by 1.7 percent per year in the median quintile and 0.8 percent per year in the largest quintile.

On average, then, institutions have been selling smaller stocks and buying larger stocks. This corresponds nicely with the trade intensity histograms in Figure 3, which show that the smallest stocks tend to have larger-size sales than buys, while the largest stocks have larger-size buys than sells. If institutions more likely trade in large sizes, we would expect this pattern. The behavior of mid-cap stocks is however anomalous in that these stocks have larger-size sales than buys despite their growth in institutional ownership.

3. Predicting institutional ownership

3.1. Regression methodology

In the market microstructure literature, institutional trading behavior has generally been identified using a cutoff rule. Trades above an upper cutoff size are classified as institutional, trades below a lower cutoff size are classified as individual, and intermediate-size trades are unclassified. Lee and Radhakrishna (2000) evaluate alternative cutoff rules using the TORQ dataset. As an example of their findings, they recommend an upper cutoff of $20,000 in small stocks. 84 percent of individual investors' trades are smaller than this, and the likelihood of finding an individual initiated trade larger than this size is 2 percent.

Our methodology refines the idea of using an optimally chosen cutoff rule. We match the TAQ data at a variety of trade sizes to the Spectrum data for a broad cross-section of stocks, over our entire sample period. That is, we use the intra-quarter tape to predict institutional ownership at the end of the quarter. Our predictive regression combines information from various trade size bins in the way that best explains the quarterly changes in institutional ownership identified in Spectrum.

We begin with extremely simple regressions that ignore the information in trade sizes. Writing Y_{it} for the share of firm i that is owned by institutions at the end of quarter t, U_{it} for unclassifiable trading volume, B_{it} for total buy volume, and S_{it} for total sell volume in stock i during quarter t, we estimate

$$\Delta Y_{it} = \alpha + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_B B_{it} + \beta_S S_{it} + \varepsilon_{it}. \tag{3.1}$$

This regression tells us how much of the variation in institutional ownership can be explained simply by the upward drift in institutional ownership of all stocks (the intercept coefficient α), mean-reversion in the institutional share for particular stocks (the autoregressive coefficient ϕ), and the total unclassifiable, buy, and sell volumes during the quarter (the coefficients β_U, β_B, and β_S). An even simpler variant of this regression restricts the coefficients on buy and sell volume to be equal and opposite, so that the explanatory variable becomes net order flow $F_{it} = B_{it} - S_{it}$ and we estimate

$$\Delta Y_{it} = \alpha + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_F F_{it} + \varepsilon_{it}. \tag{3.2}$$

We also consider variants of these regressions in which the intercept α is replaced by time dummies that soak up time-series variation in the institutional share of the stock market as a whole. In this case the remaining coefficients are identified purely by cross-sectional variation in institutional ownership, and changes in this cross-sectional variation over time.

Table III reports estimates of equation (3.1) in the top panel, and equation (3.2) in the bottom panel. Within each panel, column A restricts the lagged dependent variable and unclassifiable volume to have zero coefficients, column B restricts only the lagged dependent variable, and column C is unrestricted. Columns D, E, and F repeat these specifications including time dummies rather than an intercept. The results are remarkably consistent across all specifications. On average, buy volume gets a coefficient of about 0.36 and sell volume gets a coefficient of about –0.44. This suggests that institutions tend to use market orders, buying at the ask and selling at the bid or buying on upticks and selling on downticks, so that their orders dominate classifiable volume. The larger absolute value of the sell coefficient tells us that institutions are particularly likely to behave in this way when they are selling. The autoregressive coefficient is negative, and small but precisely estimated, telling us that there is statistically detectable mean-reversion in institutional ownership.

The coefficient on unclassifiable volume is small and only marginally significant when buys and sells are included separately in equation (3.1), but it becomes significantly negative when buys and sells are restricted to have equal and opposite coefficients in equation (3.2). To understand this, note that a stock with an equal buy and sell volume is predicted to have declining institutional ownership in the top panel of Table III. The net flow regression in the bottom panel cannot capture this effect through the net flow variable, which is identically zero if buy and sell volume are equal. Instead, it captures the effect through a negative coefficient on unclassifiable volume, which is correlated with total volume.

Table IV repeats the unrestricted regressions, incorporating time dummies, for the five quintiles of market capitalization. The main result here is that the coefficients on buys, sells, and net flows are strongly increasing in market capitalization. Evidently trading volume is more informative about institutional ownership in large firms than in small firms. The explanatory power of these regressions is U-shaped in market capitalization, somewhat above 7 percent for the smallest and largest firms and just above 5 percent for the median size firms. This is consistent with the fact, reported in Table II, that institutional ownership has the greatest cross-sectional volatility in mid-cap firms.

3.2. The information in trade size

The above summary regressions ignore the information contained in trade size. We now generalize our specification to allow separate coefficients on buy and sell volume in each trade size bin:

$$\Delta Y_{it} = \alpha + \phi Y_{i,t-1} + \beta_U U_{it} + \sum_Z \beta_{BZ} B_{Zit} + \sum_Z \beta_{SZ} S_{Zit} + \varepsilon_{it}, \tag{3.3}$$

where Z indexes trade size. In the case where we use net flows rather than separate buys and sells, the regression becomes

$$\Delta Y_{it} = \alpha + \phi Y_{i,t-1} + \beta_U U_{it} + \sum_Z \beta_{FZ} F_{Zit} + \varepsilon_{it}. \tag{3.4}$$

Table V estimates equation (3.4) separately for each quintile of market capitalization, replacing the intercept α with time dummies. It is immediately apparent that the coefficients tend to be negative for smaller trades and positive for larger trades, consistent with the intuition that order flow in small sizes reflects individual buying while order flow in large sizes reflects institutional buying. There is however an interesting exception to this pattern. Extremely small trades of less than \$2,000 have a significantly positive coefficient in the smallest three quintiles of firms, and in all quintiles have a coefficient that is much larger than that for somewhat larger trades. This suggests either that institutions break trades into extremely small sizes when they are "stealth trading" (trying to conceal their activity from the market), or that institutions are likely to engage in "scrum trades" to round off extremely small equity positions.[4] A third possibility is that these trades are in fact by individuals, but they are correlated with unobserved variables (such as news events). This could generate unclassifiable volume from institutions in a direction consistent with small

[4] Chakravarty (2001) presents an in-depth analysis of stealth trading (defined, consistently with Barclay and Warner (1993) as the trading of informed traders that attempt to pass undetected by the market maker). He shows that stealth trading (i.e., trading that is disproportionately likely to be associated with large price changes) occurs primarily via medium-sized trades by institutions of 500-9,999 shares. This runs contrary to our result here.

trades.

These results are illustrated graphically in Figures 4 and 5. These figures standardize the net flow coefficients, subtracting their mean and dividing by their standard deviation so that the set of coefficients has mean zero and standard deviation one. The standardized coefficients are then plotted against trade size. Figure 4 shows the net flow coefficients for the median quintile together with a confidence interval two standard errors above and below the coefficients, while Figure 5 shows the net flow coefficients for the smallest, median, and largest quintiles. In all cases the trough for trade sizes between $2,000 and $30,000 is clearly visible.

Figures 6 through 9 repeat the graphical presentation of coefficients for the case where buys and sells are included separately in the trade-size regression. The figures show a trough and subsequent hump for buy coefficients, and a hump and subsequent trough for sell coefficients, consistent with the net flow results.

The information in trade sizes adds considerable explanatory power to our regressions. Comparing the second panel in Table IV with Table V, the R^2 statistics increase from 7.1 percent to 8.7 percent in the smallest quintile, from 5 percent to 11.4 percent in the median quintile, and from 7.7 percent to 10.8 percent in the largest quintile. The corresponding numbers for the trade-size regressions incorporating buys and sells separately are: R^2 statistics increase from 7.2 to 10.9 percent in the smallest quintile, from 5.3 percent to 12.5 percent in the median quintile, and from 7.9 percent to 11.2 percent in the largest quintile. Of course, these R^2 statistics remain fairly modest,[5] but it should not be surprising that institutional trading activity is hard to predict given the incentives that institutions have to conceal their activity and the considerable overlap between the trade sizes that may be used by wealthy individuals and by smaller institutions.

Table VI shows that our regressions are a considerable improvement over the naive cutoff approach used in the previous market microstructure literature. The cutoff model can be thought of as a restricted regression where buys in sizes above the upper cutoff get a

[5] Note that these R^2 statistics are computed after time specific fixed effects are removed.

coefficient of plus one, buys in sizes below the lower cutoff get a coefficient of minus one, and buys in intermediate sizes get a coefficient of zero. We estimate variants of this regression in Table VI, allowing greater flexibility in successive specifications. When the coefficient restrictions imposed by the naive approach are imposed, we find that the R^2 statistic in most cases is negative. In fact, the R^2 statistic given the restrictions on the flows above and below the cutoffs is never positive for the two smallest size quintiles, and maximized at 2.5 percent, 3.5 percent and 3.8 percent respectively for the median, fourth and largest quintiles respectively. When we allow flows above and below the cutoffs to have free coefficients, the R^2 statistics of the regressions increase substantially but are still well below those of our freely estimated regressions.

3.3. Smoothing the effect of trade size

One concern about the specifications (3.3) and (3.4) is that they require the separate estimation of a large number of coefficients. This is particularly troublesome for small stocks, where large trades are extremely rare; the coefficients on large-size order flow may just reflect a few unusual trades. One way to handle this problem is to estimate a smooth function relating the buy, sell, or net flow coefficients to the trade size of the bin. We have considered polynomials in trade size, and also the exponential function suggested by Nelson and Siegel (1987) to model yield curves. We find that the Nelson and Siegel method is well able to capture the shape suggested by our unrestricted specifications. For the net flow equation, the method requires estimating a function $\beta(Z)$ that varies with trade size Z, and is of the form:

$$\beta(Z) = b_0 + (b_1 + b_2)\,[1 - e^{-Z/\tau}]\frac{\tau}{Z} - b_2 e^{-Z/\tau}. \tag{3.5}$$

Here b_0, b_1, b_2, and τ are parameters to be estimated. The parameter τ is a constant that controls the speed at which the bin-specific regression coefficients decay to zero. We estimate the function by nonlinear least squares, picking different starting values of τ, to select the function that maximizes the R^2 statistic:

$$\Delta Y_{it} = \alpha + \phi Y_{i,t-1} + \beta_U U_{it} + b_0 \sum_Z F_{Zit} + b_1 \sum_Z g_1(Z) F_{Zit} + b_2 \sum_Z g_2(Z) F_{Zit} + \varepsilon_{it}, \quad (3.6)$$

where $g_1(Z) = \frac{\tau}{Z}(1 - e^{-Z/\tau})$ and $g_2(Z) = \frac{\tau}{Z}(1 - e^{-Z/\tau}) - e^{-Z/\tau}$.

Figure 10 presents the trade-size coefficients implied by estimating (3.6). The pattern of coefficients in Figure 5 is accentuated and clarified. As before, the figure standardizes the net flow coefficients, subtracting their mean and dividing by their standard deviation so that the set of coefficients has mean zero and standard deviation one. The R^2 statistics from estimates of equation (3.6) are 8 percent for the smallest quintile of stocks, 9.9 percent for the median quintile, and 9.3 percent for the largest quintile. Figures 11 and 12 present buy and sell coefficients estimated using an analogous Nelson-Siegel specification. Again, the shapes suggested by Figures 7 and 9 appear in these figures.

The parsimony of equation (3.6) makes it relatively straightforward to explore changes in the functional form over time, as well as interactions with firm characteristics and market conditions. We hope to report the results of such explorations in the next draft of this paper.

4. Conclusion

This paper has presented a technique for predicting quarterly institutional ownership using the "tape", the publicly available record of all trades and quotes within the quarter. The technique can be used to track high-frequency institutional trading in a large cross-section of stocks. In future research we plan to use this approach to measure patterns of institutional behavior around earnings announcements, stock splits, and other corporate actions.

The results of this paper shed light on the trading behavior of institutions. Total classifiable buy volume predicts increasing institutional ownership and total sell volume predicts decreasing institutional ownership. That is, institutions tend to buy at the ask and sell at the bid, or buy on upticks and sell on downticks, suggesting that they demand liquidity rather than provide it. The coefficient on total sell volume is larger in absolute value than the coefficient on total buy volume, suggesting that institutions are particularly likely to demand liquidity when they sell. All these patterns are more pronounced in large stocks than in small stocks.

Classifying transactions by their size adds considerable explanatory power to our regressions. Buy volume in sizes between \$2,000 and \$30,000 is associated with decreasing institutional ownership, while buy volume in larger sizes predicts increasing institutional ownership. Interestingly, extremely small buys below \$2,000 also predict increasing institutional ownership, suggesting that institutions use these trades to conceal their activity or to round small positions up or down. All these patterns are reversed for sell volume, and are remarkably consistent across firm sizes.

5. Appendices

5.1. Appendix 1: Buy-Sell Classification

TAQ does not classify transactions as either buys or sells. To classify the direction of each trade, we use a matching algorithm suggested by Lee and Ready (1991). This algorithm looks at the trade price relative to quotes to determine whether a transaction is a buy or sell. The method works by matching trades to pre-existing quotes, based on time stamps. More precisely, we inspect quotes lagged by at least five seconds to avoid problems of stale reporting of quotes. If the trade price lies between the quote midpoint and the upper (lower) quote, the trade is classified as a buy (sell). If the trade price lies at the midpoint of the quotes, we use a tick test, which classifies trades that occur on an uptick as buys, and those on a downtick as sells. If the trade price lies at the midpoint of the quotes and the transactions price has not moved since the previous trade (trade occurs on a "zerotick"), Lee and Ready suggest classifying the trade based on the last recorded move in the transactions price. If the last recorded trade was classified as a buy (sell), then the zerotick trade is classified as a buy (sell). From Lee and Ready, trade-to-quote matching can be accomplished in 75.7% of trades, while tick tests are required in 23.8% of cases. The remaining trades take place outside the quoted spread.

The analysis in Lee and Radhakrishna (2000) evaluates the effectiveness of the Lee and Ready matching algorithm, using the TORQ database, which has buy-sell classified, institutional-individual identified data for 144 stocks over a 3 month period. They find that after removing trades with potentially ambiguous classifications (such as trades that are batched or split up during execution), the buy/sell classification algorithm is 93 percent effective. In particular, they find that the accuracy is highest (at 98 percent) when trade-to-quote matching can be accomplished, lower (at 76 percent) for those trades that have to be classified using a tick test, and lowest (at 60 percent) for those trades classified using a zerotick test.

We eliminate this last source of variability in our data by terming as unclassifiable those trades for which a zerotick test is required. We further identify as unclassifiable all trades that occur in the first half hour of trading (since these come from the opening auction) as well as any trade that is reported as cancelled, batched or split up in execution. This last category of trades is identified as unclassifiable since we use trade size as one important input into our prediction of institutional ownership. A trade that is reported as being batched or split up cannot be unambiguously classified in terms of its size. We aggregate all unclassifiable trades together, and use the bin of unclassifiable trades as an additional input into our prediction exercise.

5.2. Appendix 2: Spectrum Institutional Ownership Data

A 1978 amendment to the Securities and Exchange Act of 1934 required all institutions with greater than $100 million of securities under discretionary management to report their holdings to the SEC. Holdings are reported quarterly on the SEC's form 13F, where all common-stock positions greater than 10,000 shares or $200,000 must be disclosed. These reports are available in electronic form back to 1980 from CDA/Spectrum, a firm hired by the SEC to process the 13F filings. Our data include the quarterly reports from the first quarter of 1993 to the final quarter of 2001. Throughout this paper, we use the term institution to refer to an institution that files a 13F. On the 13F, each manager must report all securities over which they exercise sole or shared investment discretion. In cases where investment discretion is shared by more than one institution, care is taken to prevent double counting.

Our Spectrum data have been extensively cleaned by Kovtunenko and Sosner (2003). They first identify all inconsistent records, those for which the number of shares held by an institution in a particular stock at the end of quarter $t-1$ is not equal to the number of shares held at the end of quarter t minus the reported net change in shares since the prior quarter. They assume that the holdings data are correct for such observations, rather than the reported change data.

They proceed to fill in missing records, using the general rule that if a stock has a return on CRSP but does not have reported Spectrum holdings in a given quarter, holdings are set to zero. For the missing records inconsistent with this assumption (those for which holdings at the end of quarter t are above the reported net change from previous quarter holdings), they fill in the holdings for the end of quarter $t-1$ as split-adjusted holdings in period t less the reported net change in holdings.

The Spectrum 13F holdings file contains three columns: date, CUSIP code, identifier for the institution, and number of shares held in that stock by that institution on that date. All dates are end-of-quarter (March 31, June 30, September 30, or December 31). For each CUSIP and date we simply sum up the shares held by all institutions in the sample to get total institutional holdings of the security at the end of that quarter.

References

[1] Arnold, T. and M. Lipson 1997, "Tick size and Limit Order Executions: An Examination of Stock Splits", Unpublished working paper.

[2] Chakravarty, S. 2001, "Stealth-Trading: Which Traders' Trades Move Stock Prices?" *Journal of Financial Economics,* 61, 2, 289–307

[3] Chen, J., Hong, H. and J. Stein, 2002, "Breadth of Ownership and Stock Returns," *Journal of Financial Economics*, 66, 171-205.

[4] Chen, H., N. Jegadeesh, and R. Wermers, 2000, "The value of active mutual fund management: an examination of the stockholdings and trades of fund managers," *Journal of Financial and Quantitative Analysis,* 35, 343-368.

[5] Cohen, R., P. Gompers, and T. Vuolteenaho, 2002, "Who underreacts to cashflow news? Evidence from trading between individuals and institutions," *Journal of Financial Economics,* 66, 2-3, pp.409-462.

[6] Daniel, K. D., M. Grinblatt, S. Titman, and R. Wermers, 1997, "Measuring mutual fund performance with characteristic-based benchmarks," *Journal of Finance,* 52, 1035-1058.

[7] Daves, P., J. Wansley, and R. Zhang, 2003, "Reported Trading Volume on the NYSE and the Nasdaq," working paper, University of Tennessee.

[8] Easley, D., O'Hara, M. and G. Saar, 2001, "How Stock Splits Affect Trading: A Microstructure Approach," *Journal of Financial and Quantitative Analysis*, 36, 25-51.

[9] Ellis, K., Michaely, R. and M. O'Hara, 2000, "The accuracy of trade classification rules: Evidence from Nasdaq," *Journal of Financial and Quantitative Analysis.*

[10] Finucane, Thomas, J., 2000, "A direct test of methods for inferring trade direction from intra-day data" *Journal of Financial and Quantitative Analysis*, 35, 4, 553-577.

[11] Gompers, Paul A., and Andrew Metrick, 2001, "Institutional investors and equity prices," *Quarterly Journal of Economics* 116, 229-260.

[12] Greene, Jason, 1995, "The impact of limit order executions on trading costs in NYSE stocks: an empirical examination" Indiana University Working Paper

[13] Grinblatt, Mark and Matti Keloharju, 2000a, "The investment behavior and performance of various investor-types: A study of Finland's unique data set," *Journal of Financial Economics* 55, 43-67.

[14] Grinblatt, Mark and Matti Keloharju, 2000b, "What makes investors trade?," *Journal of Finance*, 56, 589-616.

[15] Kovtunenko, Boris and Nathan Sosner, 2003, "Sources of Institutional Performance," unpublished working paper, Harvard University.

[16] Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, 1992, "The Impact of Institutional Trading on Stock Prices," *Journal of Financial Economics*, 32, 1, pp. 23-43

[17] Lee, Charles M. C. and Mark J. Ready, 1991, "Inferring trade direction from intraday data" *Journal of Finance* 46, 733-746.

[18] Lee, Charles M. C. and Balkrishna Radhakrishna, 2000,. "Inferring Investor Behavior: Evidence from TORQ Data," *Journal of Financial Markets* 3(2): 83-111.

[19] Nelson, Charles R. and Andrew F. Siegel, 1987, "Parsimonious Modeling of Yield Curves," *Journal of Business* 60, 473-489.

[20] Nofsinger, John R. and Richard W. Sias, 1999, "Herding and feedback trading by institutional and individual investors," *Journal of Finance* 54, 2263-2295.

[21] Vuolteenaho, Tuomo, 2002, "What drives firm-level stock returns?," *Journal of Finance*, 57 (1), pp. 233-264.

[22] Wermers, Russ, 1999, "Mutual fund herding and the impact on stock prices," *Journal of Finance* 54, 581-622.

[23] Wermers, Russ, 2000, "Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses," *Journal of Finance* 55, 1655-1695.

Table I: Summary Statistics for Firm Size Quintiles

Table I presents means, medians and standard deviations for the TAQ and Spectrum variables in our specifications. All data are independently winsorized at the 2.5 standard deviation level. The variables are in sequence, the total buyer initiated orders in TAQ classified by the Lee and Ready algorithm; the total seller initiated orders, similarly classified; the total unclassifiable volume (those transacted in the opening auction, reported as cancelled, or unclassifiable as a buy or a sell by the LR algorithm); the total volume (the sum of the previous three variables); the net order imbalance (total classifiable buys less total classifiable sells); and finally, the change in quarterly 13-F institutional ownership as reported in the Spectrum dataset as a fraction of CRSP shares outstanding. All TAQ variables are normalized by daily shares outstanding as reported in CRSP, and then summed up to the quarterly frequency. All summary statistics are presented as annualized percentages (standard deviations are annualized under the assumption that quarterly observations are iid). The columns report these summary statistics first for all firms, and then for firm size quintiles, where firms are sorted quarterly by market capitalization (size).

	All	Small	Q2	Q3	Q4	Large
Mean						
TAQ Total Buys	31.83	20.88	27.42	33.56	39.20	38.04
TAQ Total Sells	30.99	23.13	28.58	33.00	36.45	33.75
TAQ Unclassifiable	15.39	9.66	13.21	15.94	18.85	19.29
TAQ Total Volume	78.31	53.82	69.28	82.62	94.61	91.14
TAQ Net Imbalance	0.96	-2.13	-1.08	0.63	2.87	4.49
Spectrum Change	0.60	-1.31	0.29	1.73	1.49	0.77
Median						
TAQ Total Buys	23.84	13.72	18.76	24.85	31.40	30.58
TAQ Total Sells	23.90	15.80	20.63	25.58	29.90	27.41
TAQ Unclassifiable	11.57	5.70	8.73	11.47	15.04	15.81
TAQ Total Volume	60.42	36.43	49.26	63.03	77.00	74.35
TAQ Net Imbalance	0.55	-1.22	-0.62	0.15	1.62	3.09
Spectrum Change	0.43	-0.03	0.41	1.65	1.35	0.98
Standard Deviation						
TAQ Total Buys	13.48	11.00	13.04	13.84	14.23	12.82
TAQ Total Sells	12.40	11.32	12.55	12.81	12.83	11.30
TAQ Unclassifiable	6.79	5.69	6.69	7.00	7.03	6.21
TAQ Total Volume	31.68	27.09	31.27	32.62	33.06	29.46
TAQ Net Imbalance	5.07	4.87	5.07	5.22	5.15	4.23
Spectrum Change	8.94	7.48	9.40	9.87	9.59	8.03

Table II: Summary Statistics for Bins and Firm Size Quintiles

Table II presents medians (top panel) and standard deviations (bottom panel) for total TAQ buys + sells classified by the Lee and Ready algorithm (normalized by firm shares outstanding). All data are winsorized at the 2.5 standard deviation level. All summary statistics are reported in annualized percentage terms (standard deviations are annualized under the assumption that quarterly observations are iid).

Median	**Small**	**Q2**	**Q3**	**Q4**	**Large**
Buys + Sells					
0-2000	2.126	0.680	0.163	0.000	0.000
2000-3000	1.727	0.728	0.409	0.166	0.000
3000-5000	3.438	1.675	1.004	0.700	0.319
5000-7000	2.597	1.738	1.045	0.691	0.405
7000-9000	1.943	1.623	1.051	0.713	0.385
9000-10000	0.793	0.710	0.512	0.330	0.139
10000-20000	4.902	6.038	5.025	3.555	1.954
20000-30000	2.056	3.351	3.526	3.070	1.831
30000-50000	1.784	3.934	4.584	4.692	3.215
50000-70000	0.544	2.378	3.085	3.353	2.630
70000-90000	0.000	1.628	2.352	2.737	2.194
90000-100000	0.000	0.586	0.998	1.180	0.988
100000-200000	0.000	3.562	6.626	8.567	7.630
200000-300000	0.000	1.329	3.411	5.059	5.064
300000-500000	0.000	1.139	3.623	5.941	6.554
500000-700000	0.000	0.000	1.853	3.452	4.116
700000-900000	0.000	0.000	1.067	2.237	2.814
900000-1000000	0.000	0.000	0.000	0.792	1.087
>1000000	0.000	0.000	4.018	9.543	14.178

Standard Deviation	**Small**	**Q2**	**Q3**	**Q4**	**Large**
Buys + Sells					
0-2000	1.905	1.002	0.529	0.317	0.134
2000-3000	1.095	0.744	0.459	0.324	0.197
3000-5000	1.861	1.444	0.910	0.610	0.382
5000-7000	1.575	1.348	0.910	0.610	0.404
7000-9000	1.336	1.228	0.892	0.606	0.376
9000-10000	0.658	0.620	0.477	0.332	0.224
10000-20000	3.554	3.610	3.009	2.293	1.421
20000-30000	2.013	2.243	2.080	1.835	1.245
30000-50000	2.216	2.720	2.679	2.514	1.903
50000-70000	1.317	1.757	1.799	1.753	1.470
70000-90000	0.940	1.329	1.399	1.372	1.155
90000-100000	0.427	0.606	0.642	0.635	0.542
100000-200000	1.981	3.193	3.730	3.840	3.291
200000-300000	1.047	1.736	2.214	2.395	2.132
300000-500000	1.188	1.942	2.600	2.925	2.699
500000-700000	0.705	1.178	1.615	1.832	1.786
700000-900000	0.467	0.861	1.168	1.329	1.296
900000-1000000	0.134	0.385	0.530	0.582	0.570
> 1000000	1.949	3.481	4.655	5.791	5.865

Table III: Regression Specifications on Total Buys, Sells and Net Flows

Table III presents estimates of several specifications, in which the dependent variable is the change in Spectrum institutional ownership as a fraction of shares outstanding. The first panel below presents the independent variables in rows: an intercept, the lagged level of Spectrum institutional ownership as a percentage of the shares outstanding of the firm, the total unclassifiable volume in TAQ, total buyer initiated trades and total seller initiated trades. The second panel uses the same first three independent variables, but uses total net flows (total buys less total sells) as the fourth independent variable. Different specifications use different combinations of these independent variables. Specifications D-F are the same as specifications A-C, except that they incorporate quarter-specific time dummy variables. White corrected t-statistics are reported in parentheses below the coefficients.

	A	B	C	D	E	F
Intercept	0.007	0.007	0.012			
	(23.669)	(23.588)	(36.240)			
Lagged Spectrum Level			-0.016			-0.015
			-(20.386)			-(20.174)
TAQ Unclassifiable		0.046	0.045		0.011	0.009
		(3.213)	(3.160)		(0.669)	(0.554)
TAQ Total Buys	0.348	0.338	0.360	0.347	0.344	0.367
	(32.879)	(29.983)	(32.022)	(32.725)	(29.965)	(31.981)
TAQ Total Sells	-0.429	-0.441	-0.438	-0.422	-0.425	-0.423
	-(36.422)	-(36.355)	-(36.302)	-(35.867)	-(34.661)	-(34.611)
R-Squared	0.040	0.041	0.047	0.039	0.039	0.045
N	66805	66805	66805	66805	66805	66805
N(Firms)	3402	3402	3402	3402	3402	3402
Time Dummies?	No	No	No	Yes	Yes	Yes

	A	B	C	D	E	F
Intercept	0.001	0.005	0.011			
	(4.421)	(19.437)	(35.136)			
Lagged Spectrum Level			-0.017			-0.016
			-(23.283)			-(22.082)
TAQ Unclassifiable		-0.113	-0.071		-0.115	-0.075
		-(14.614)	-(8.711)		-(14.854)	-(9.137)
TAQ Net Flows	0.337	0.371	0.387	0.338	0.373	0.388
	(31.804)	(33.924)	(35.635)	(31.910)	(34.108)	(35.687)
R-Squared	0.029	0.036	0.045	0.030	0.037	0.045
N	66805	66805	66805	66805	66805	66805
N(Firms)	3402	3402	3402	3402	3402	3402
Time Dummies?	No	No	No	Yes	Yes	Yes

Table IV: Size Quintile Specific Regressions of Spectrum Change on Total TAQ Flows

This table presents estimates of specification F from Table III, estimated separately for stocks sorted into market capitalization quintiles. The dependent variable in all specifications is the change in Spectrum institutional ownership as a fraction of shares outstanding. The first panel below presents the independent variables in rows: the lagged level of Spectrum institutional ownership as a percentage of the shares outstanding of the firm, the total unclassifiable volume in TAQ, total buyer initiated trades and total seller initiated trades. The second panel uses the same first three independent variables, but uses total net flows (total buys less total sells) as the fourth independent variable. All specifications incorporate quarter-specific time dummy variables. White corrected t-statistics are reported in parentheses below the coefficients.

	Small	**Q2**	**Q3**	**Q4**	**Large**
Lagged Spectrum Level	-0.042	-0.029	-0.024	-0.024	-0.033
	-(15.350)	-(13.686)	-(11.664)	-(10.937)	-(12.584)
TAQ Unclassifiable	-0.075	0.024	0.072	0.042	0.032
	-(2.273)	(0.620)	(1.945)	(1.167)	(0.913)
TAQ Total Buys	0.164	0.204	0.330	0.461	0.557
	(6.325)	(7.636)	(12.999)	(17.957)	(20.120)
TAQ Total Sells	-0.216	-0.270	-0.431	-0.530	-0.648
	-(8.751)	-(10.392)	-(16.101)	-(18.251)	-(20.425)
R-Squared	0.072	0.041	0.053	0.062	0.079
N	13341	13370	13361	13349	13384
N(Firms)	1198	1422	1366	1194	744
Time Dummies?	Yes	Yes	Yes	Yes	Yes

	Small	**Q2**	**Q3**	**Q4**	**Large**
Lagged Spectrum Level	-0.043	-0.030	-0.026	-0.026	-0.035
	-(15.527)	-(14.301)	-(13.214)	-(12.380)	-(14.093)
TAQ Unclassifiable	-0.143	-0.078	-0.088	-0.064	-0.101
	-(8.407)	-(4.152)	-(4.668)	-(3.572)	-(5.833)
TAQ Net Flows	0.193	0.236	0.366	0.480	0.578
	(8.383)	(9.883)	(15.322)	(19.335)	(21.181)
R-Squared	0.071	0.040	0.050	0.061	0.077
N	13341	13370	13361	13349	13384
N(Firms)	1198	1422	1366	1194	744
Time Dummies?	Yes	Yes	Yes	Yes	Yes

Table V: Estimates of Spectrum-TAQ Quarterly Predictive Regression

This table presents estimates from an equation relating quarterly Spectrum institutional ownership to TAQ for different size quintiles of stocks. Here, the dependent variable is the change in quarterly 13-F institutional ownership from Spectrum (as a fraction of firm shares outstanding). In order, the dependent variables are the lagged level of the Spectrum institutional ownership fraction, the total unclassifiable volume in TAQ, and Net Flows, which are the number of shares bought less shares sold traded within dollar cutoff bins from TAQ (normalized by CRSP daily shares outstanding, and then summed up to the quarterly frequency). All specifications incorporate quarter-specific time dummy variables. White corrected t-statistics are reported below coefficients in parentheses.

	Small	Q2	Q3	Q4	Large
Lagged Spectrum Level	-0.041	-0.030	-0.027	-0.025	-0.035
	-(15.014)	-(14.347)	-(13.927)	-(12.017)	-(13.416)
Total Unclassifiable	-0.123	-0.051	-0.047	-0.065	-0.082
	-(7.219)	-(2.731)	-(2.444)	-(3.534)	-(4.528)
Net Flows					
0-2000	0.804	3.546	3.928	1.263	0.165
	(4.755)	(5.270)	(2.807)	(0.540)	(0.033)
2000-3000	-0.341	-0.334	-4.039	-2.719	-5.695
	-(1.408)	-(0.380)	-(2.664)	-(1.342)	-(1.984)
3000-5000	-0.653	-2.336	-1.833	-2.408	-2.142
	-(4.414)	-(4.409)	-(1.764)	-(1.692)	-(1.013)
5000-7000	-0.246	-1.490	-3.067	0.901	-3.787
	-(1.652)	-(3.072)	-(3.014)	(0.547)	-(1.925)
7000-9000	-0.389	-0.852	-2.162	-2.420	-1.320
	-(2.310)	-(1.805)	-(2.292)	-(1.546)	-(0.625)
9000-10000	-0.051	-1.355	-3.707	-4.902	2.187
	-(0.203)	-(1.896)	-(2.757)	-(2.204)	(0.665)
10000-20000	0.067	-0.931	-1.285	-1.927	-2.422
	(0.799)	-(4.199)	-(3.416)	-(2.784)	-(2.142)
20000-30000	0.072	0.025	-0.498	-1.073	1.209
	(0.655)	(0.105)	-(1.214)	-(1.612)	(0.934)
30000-50000	0.223	0.178	0.346	0.569	-0.179
	(2.199)	(0.979)	(1.183)	(1.262)	-(0.211)
50000-70000	0.359	0.464	0.487	0.911	1.491
	(2.917)	(2.202)	(1.478)	(1.869)	(1.997)
70000-90000	0.490	0.658	0.382	1.007	1.501
	(3.084)	(2.855)	(1.118)	(2.096)	(1.744)
90000-100000	0.918	0.977	0.701	1.825	-0.086
	(3.378)	(2.741)	(1.433)	(2.517)	-(0.077)
100000-200000	0.616	0.593	0.756	0.692	0.552
	(6.552)	(5.256)	(5.226)	(3.410)	(1.566)
200000-300000	0.313	0.888	1.027	0.990	-0.044
	(2.481)	(6.133)	(6.138)	(4.364)	-(0.122)
300000-500000	0.202	0.508	0.934	0.940	0.364
	(1.788)	(4.377)	(7.180)	(5.533)	(1.423)
500000-700000	0.165	0.269	0.669	0.728	0.996
	(0.980)	(1.920)	(4.396)	(3.756)	(3.621)
700000-900000	0.318	0.380	1.024	1.057	2.178
	(1.150)	(2.344)	(5.989)	(5.053)	(7.125)
900000-1000000	1.889	0.477	0.431	1.158	3.573

	(2.153)	(1.467)	(1.559)	(3.536)	(7.472)
> 1000000	0.163	0.184	0.231	0.267	0.309
	(2.438)	(4.322)	(6.253)	(6.758)	(6.694)
R^2	0.087	0.075	0.114	0.103	0.108
N	13341	13370	13361	13349	13384
N(Firms)	1198	1422	1366	1194	744
Time Dummies?	Yes	Yes	Yes	Yes	Yes

Table VI: Evaluating the Lee-Radhakrishna Method Using Spectrum and TAQ

This table presents R^2 statistics for various specifications of the Lee-Radhakrishna regression (that employs a single cutoff level) of the change in quarterly 13-F institutional ownership as reported in the Spectrum dataset (as a fraction of CRSP shares outstanding *O*) $\Delta(S_{i,t}/O_{i,t})$ on net quarterly flows from TAQ (normalized by CRSP daily shares outstanding, and then summed up to the quarterly frequency). Here $f_{i,t}(c)$, represents net flows *greater* than \$*c*, and $f_{i,t}(-c)$ represents net flows *less* than \$*c*. We estimate variants of the following specification: $\Delta(S_{i,t}/O_{i,t}) = \alpha + \beta_{c1} f_{i,t}(-c1) + \beta_{c2} f_{i,t}(c2) + \varepsilon_{i,t}$. The specifications in rows labeled $\alpha = \hat{\alpha}_t$ include quarter-specific time dummies. The specification is estimated separately for different size quintiles of stocks (in columns). Row headings indicate estimates of the R^2 statistic under different coefficient restrictions, for different values of the dollar cutoff levels *c1* and *c2*.

R^2	Small	Q2	Q3	Q4	Large
C1 =2,000; c2=5,000					
$\alpha = 0, \beta_{c1} = -1, \beta_{c2} = 1$	-0.101	-0.105	-0.063	-0.032	-0.057
$\alpha = \hat{\alpha}_t, \beta_{c1} = -1, \beta_{c2} = 1$	-0.109	-0.107	-0.061	-0.017	-0.002
$\alpha = 0, \beta_{c1} = \hat{\beta}_{c1}, \beta_{c2} = \hat{\beta}_{c2}$	0.024	0.020	0.033	0.046	0.037
$\alpha = \hat{\alpha}_t, \beta_{c1} = \hat{\beta}_{c1}, \beta_{c2} = \hat{\beta}_{c2}$	0.024	0.020	0.040	0.049	0.045
C1=3,000;c2=10,000					
$\alpha = 0, \beta_{c1} = -1, \beta_{c2} = 1$	-0.071	-0.084	-0.042	-0.020	-0.048
$\alpha = \hat{\alpha}_t, \beta_{c1} = -1, \beta_{c2} = 1$	-0.077	-0.084	-0.039	-0.006	0.005
$\alpha = 0, \beta_{c1} = \hat{\beta}_{c1}, \beta_{c2} = \hat{\beta}_{c2}$	0.029	0.025	0.047	0.056	0.048
$\alpha = \hat{\alpha}_t, \beta_{c1} = \hat{\beta}_{c1}, \beta_{c2} = \hat{\beta}_{c2}$	0.029	0.026	0.056	0.060	0.056
C1=3,000;c2=20,000					
$\alpha = 0, \beta_{c1} = -1, \beta_{c2} = 1$	-0.047	-0.061	-0.022	-0.005	-0.037
$\alpha = \hat{\alpha}_t, \beta_{c1} = -1, \beta_{c2} = 1$	-0.049	-0.059	-0.017	0.008	0.014
$\alpha = 0, \beta_{c1} = \hat{\beta}_{c1}, \beta_{c2} = \hat{\beta}_{c2}$	0.031	0.029	0.052	0.060	0.051
$\alpha = \hat{\alpha}_t, \beta_{c1} = \hat{\beta}_{c1}, \beta_{c2} = \hat{\beta}_{c2}$	0.033	0.030	0.061	0.064	0.059
C1=3,000;c2=50,000					
$\alpha = 0, \beta_{c1} = -1, \beta_{c2} = 1$	-0.026	-0.038	-0.001	0.015	-0.017
$\alpha = \hat{\alpha}_t, \beta_{c1} = -1, \beta_{c2} = 1$	-0.024	-0.035	0.007	0.024	0.026
$\alpha = 0, \beta_{c1} = \hat{\beta}_{c1}, \beta_{c2} = \hat{\beta}_{c2}$	0.030	0.032	0.056	0.064	0.054
$\alpha = \hat{\alpha}_t, \beta_{c1} = \hat{\beta}_{c1}, \beta_{c2} = \hat{\beta}_{c2}$	0.032	0.034	0.066	0.068	0.062
C1=5,000;c2=100,000					
$\alpha = 0, \beta_{c1} = -1, \beta_{c2} = 1$	-0.029	-0.029	0.012	0.030	0.006
$\alpha = \hat{\alpha}_t, \beta_{c1} = -1, \beta_{c2} = 1$	-0.023	-0.021	0.025	0.035	0.038
$\alpha = 0, \beta_{c1} = \hat{\beta}_{c1}, \beta_{c2} = \hat{\beta}_{c2}$	0.023	0.031	0.059	0.069	0.061
$\alpha = \hat{\alpha}_t, \beta_{c1} = \hat{\beta}_{c1}, \beta_{c2} = \hat{\beta}_{c2}$	0.027	0.034	0.074	0.073	0.068
N	13341	13370	13361	13349	13384
N(Firms)	1198	1422	1366	1194	744

Figure 1

This figure plots the evolution of the number of firms in our sample across time measured in quarters. The sample consists only of firms issuing common stock on the NYSE or AMEX exchanges. The data begin in the first quarter of 1993, and end in the final quarter of 2000.

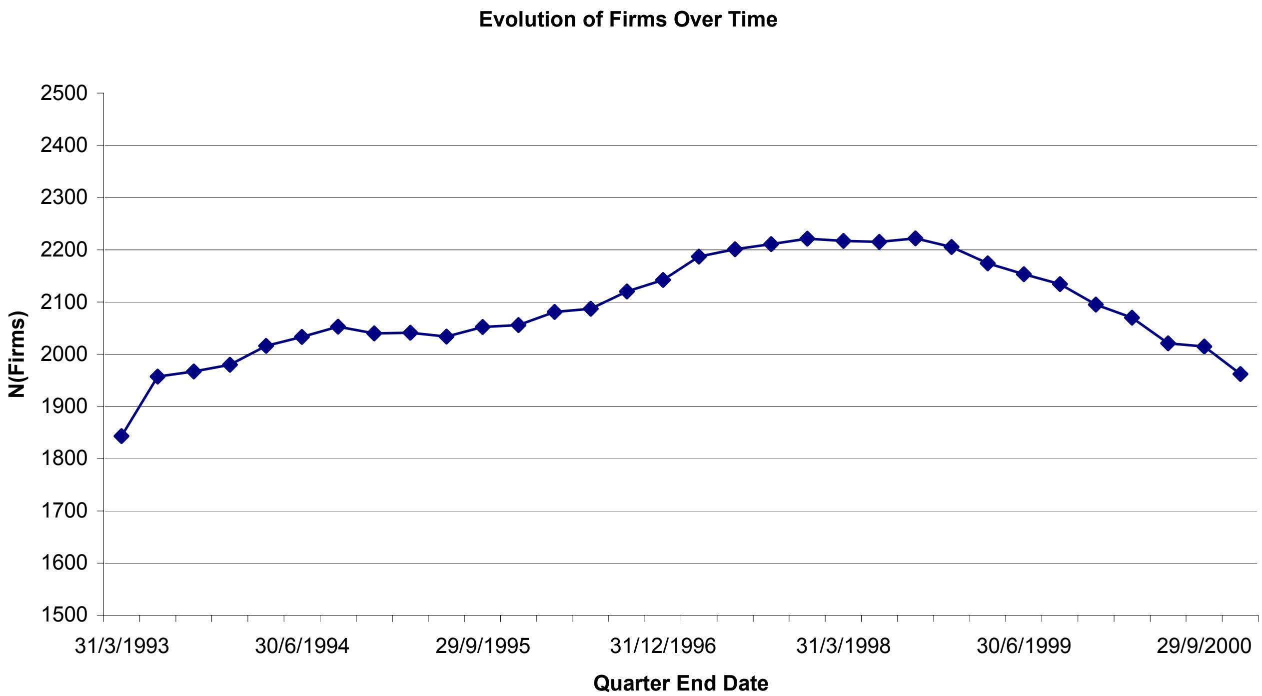


Figure 2

This figure plots the mean and standard deviation across all firms each quarter of the total volume of shares traded as a percentage of shares outstanding for each firm. The volume measure is obtained by summing all trades reported for each firm-quarter in the Transactions and Quotes (TAQ) database of the NYSE. Total shares outstanding for each firm is obtained from CRSP.

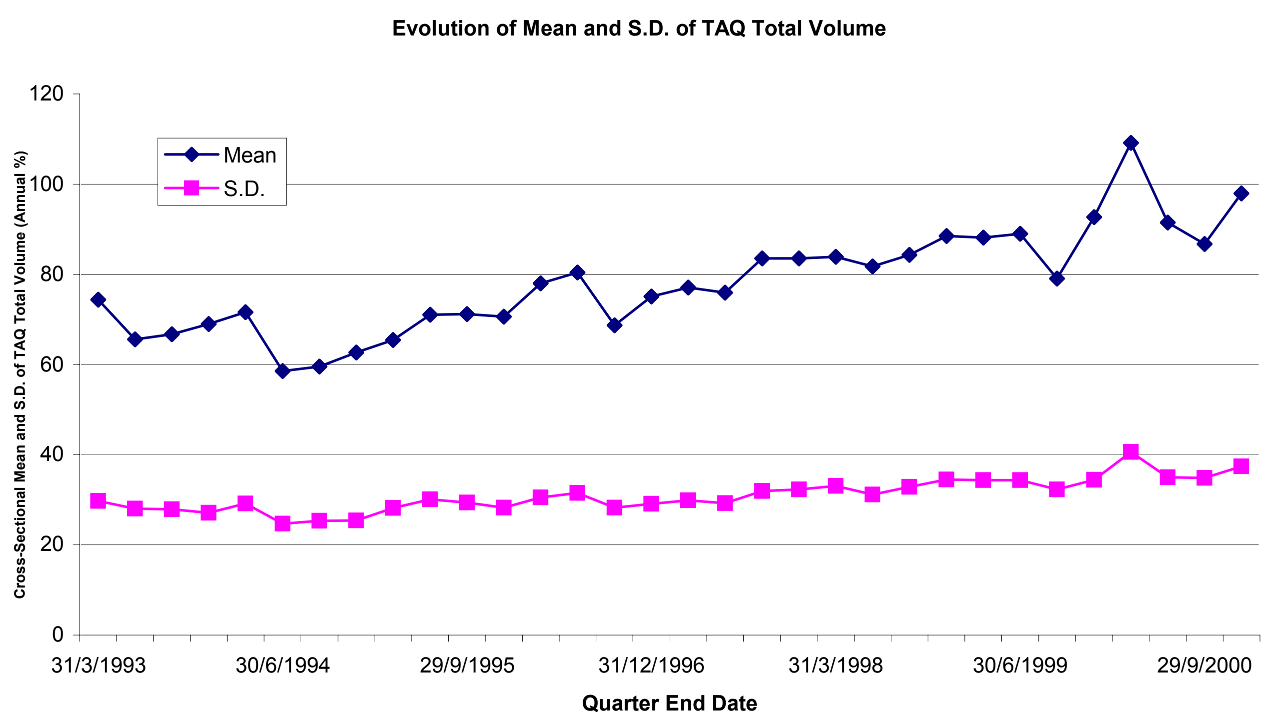
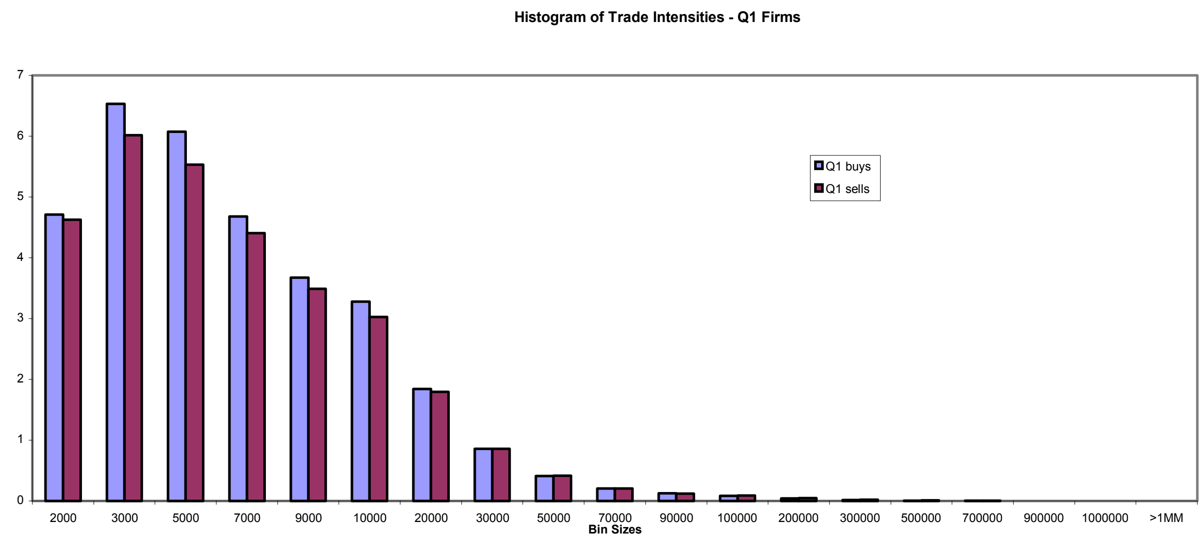
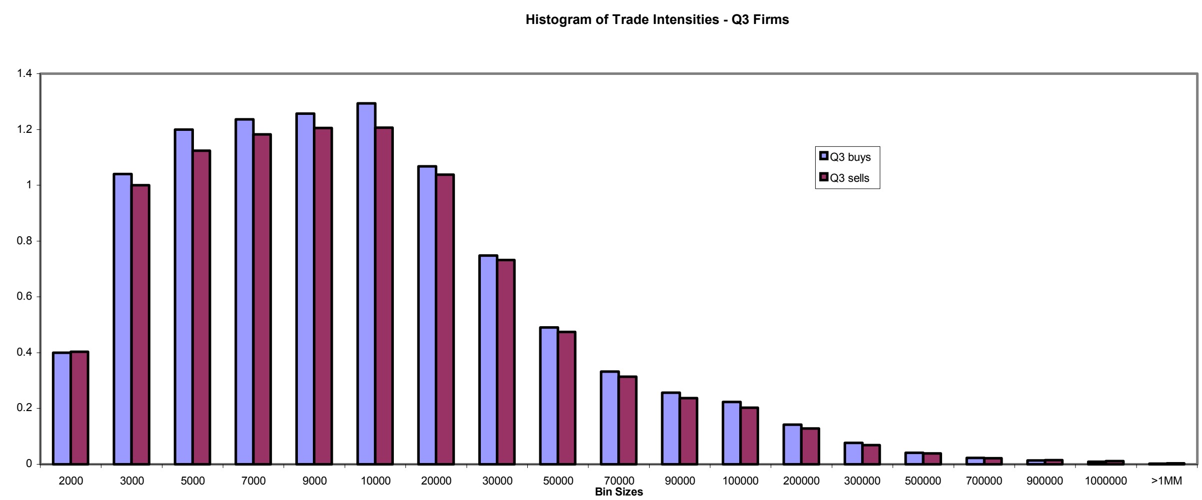


Figure 3

Figure 3 plots histograms of trade intensities (total volume as a percentage of shares outstanding in each bin divided by relative bin width), for dollar trade size bins that aggregate TAQ trades classified into buys and sells. A bin size of $5 million is assigned to the largest bin. The three panels show, in sequence, histograms for small, median and large firms sorted quarterly into quintiles based on relative market capitalization (size).





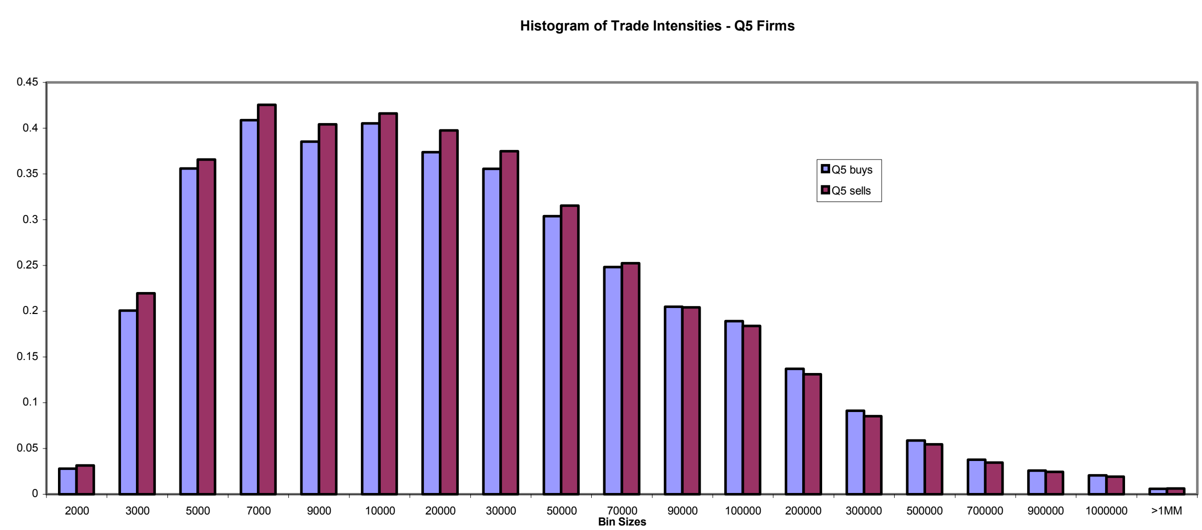


Figure 4

This figure plots the net flow coefficients for each trade size bin, for the Q3 firms in our sample. The coefficients are standardized by removing the within quintile cross-sectional mean of bin coefficients, and dividing by the cross-sectional standard deviation of bin coefficients. The dashed lines are +/- 2 OLS standard error bounds.

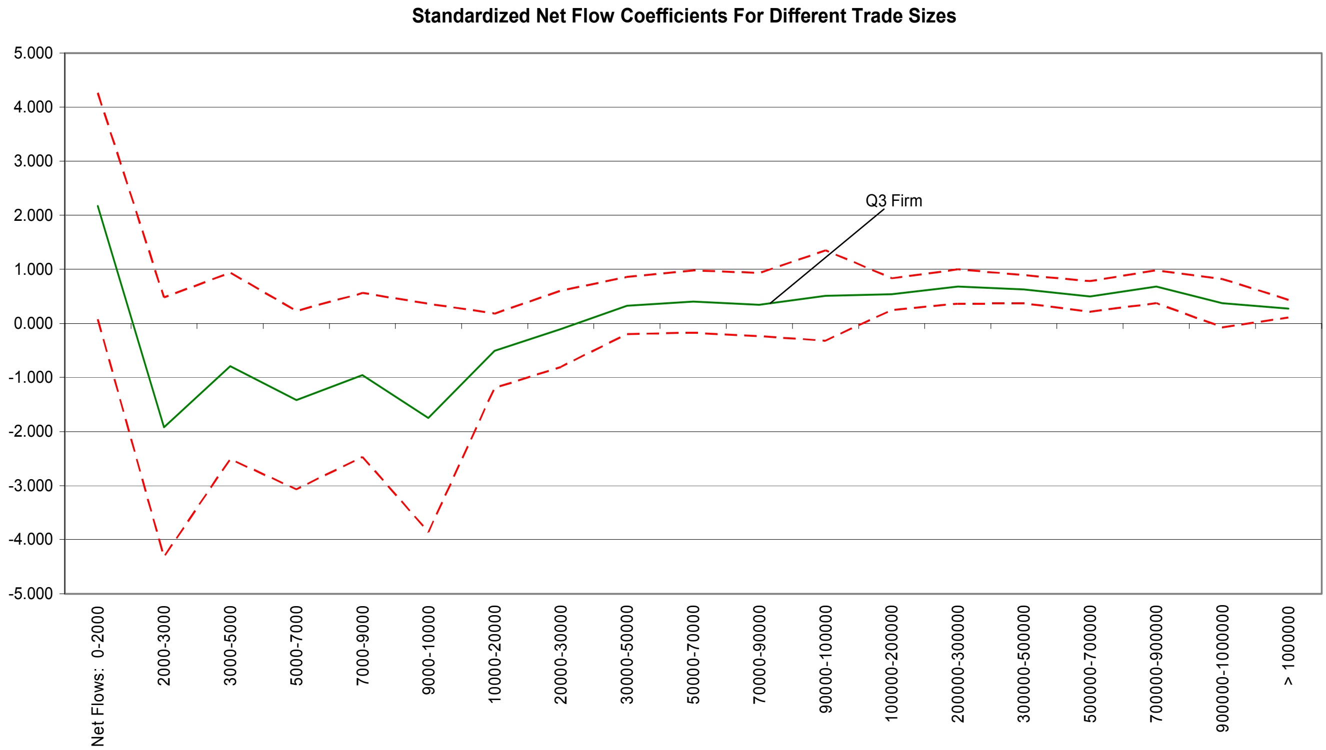


Figure 5

This figure plots the net flow coefficients for each trade size bin, for the Q1, Q3 and Q5 firms in our sample. The coefficients are standardized by removing the within quintile cross-sectional mean of bin coefficients, and dividing by the cross-sectional standard deviation of bin coefficients.

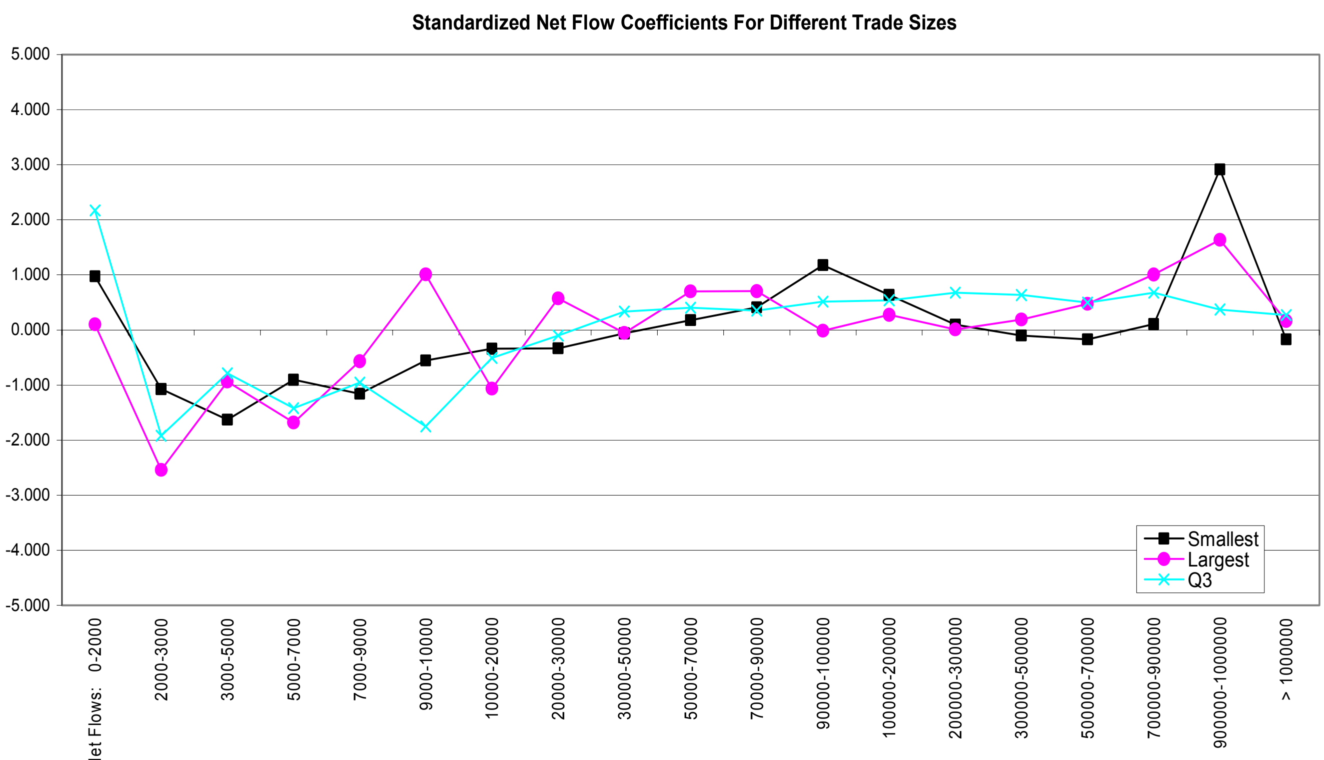


Figure 6

This figure plots the buy coefficients for each trade size bin, for the Q3 firms in our sample. The coefficients are standardized by removing the within quintile cross-sectional mean of bin coefficients, and dividing by the cross-sectional standard deviation of bin coefficients. The dashed lines are +/- 2 OLS standard error bounds.

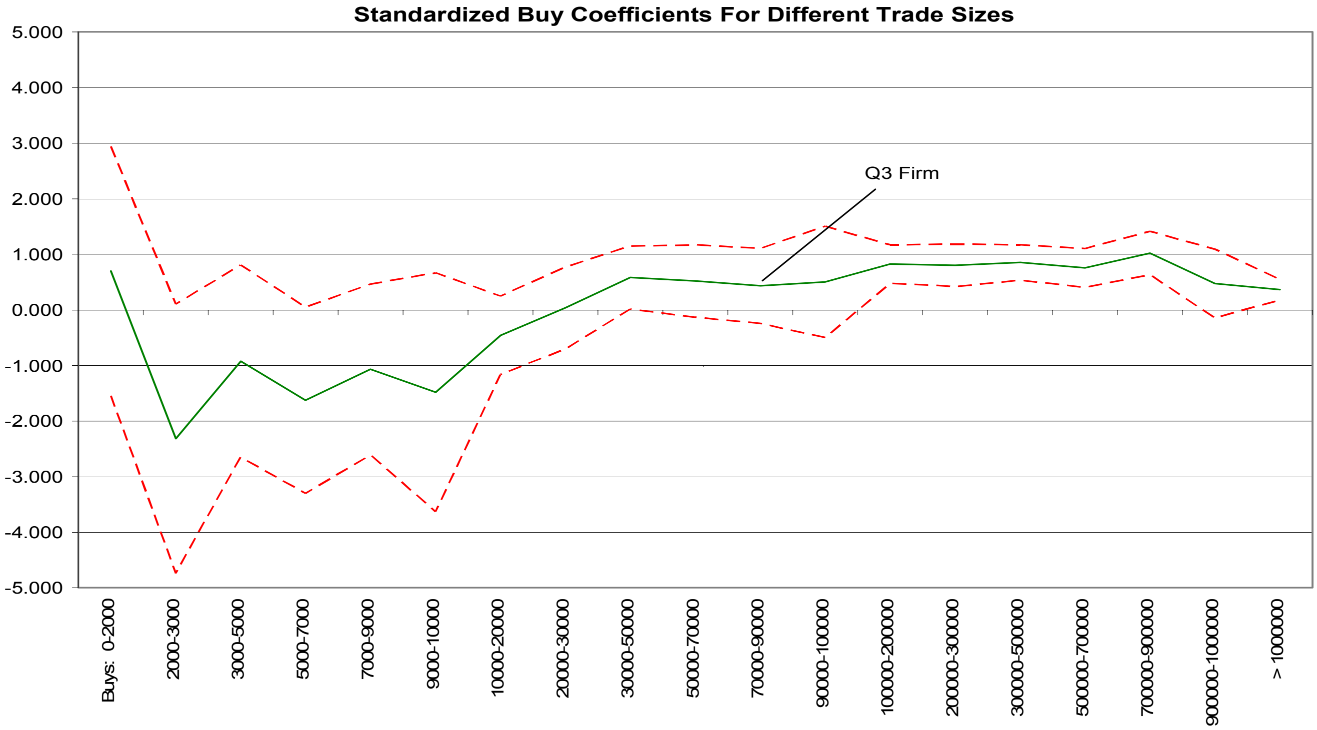


Figure 7

This figure plots the buy coefficients for each trade size bin, for the Q1, Q3 and Q5 firms in our sample. The coefficients are standardized by removing the within quintile cross-sectional mean of bin coefficients, and dividing by the cross-sectional standard deviation of bin coefficients.

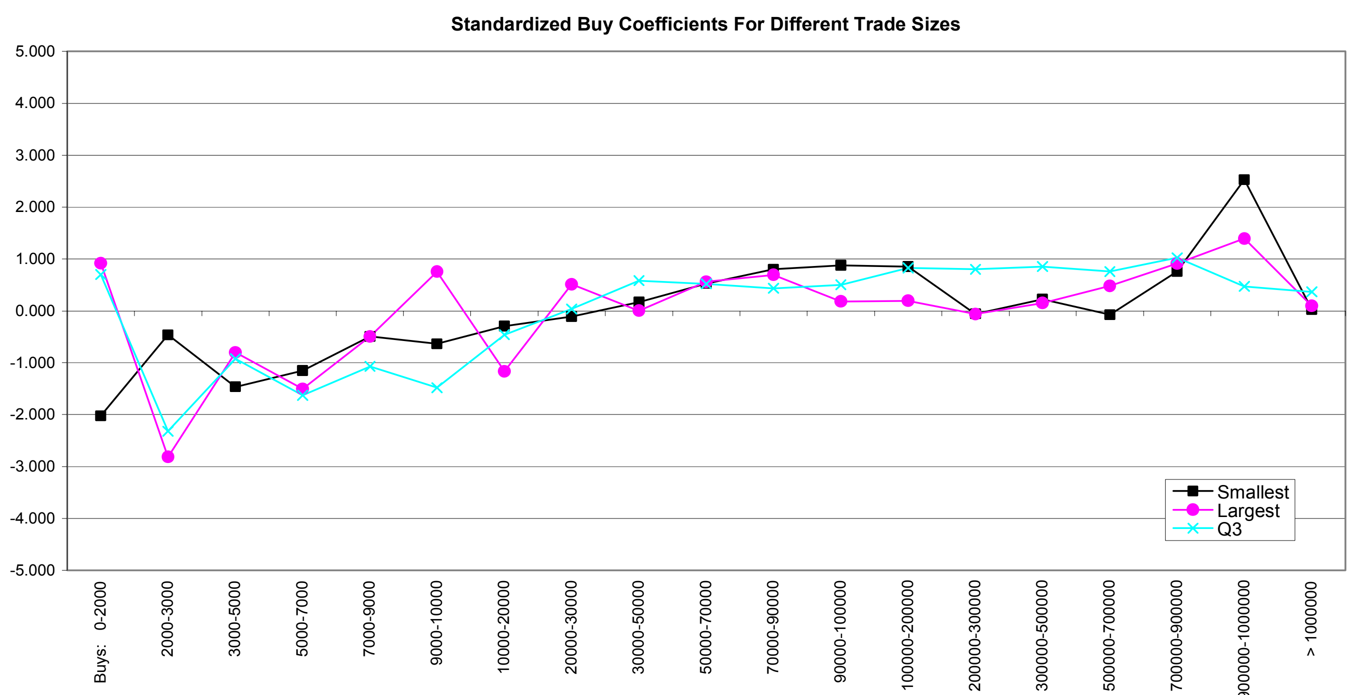


Figure 8

This figure plots the sell coefficients for each trade size bin, for the Q3 firms in our sample. The coefficients are standardized by removing the within quintile cross-sectional mean of bin coefficients, and dividing by the cross-sectional standard deviation of bin coefficients. The dashed lines are +/- 2 OLS standard error bounds.

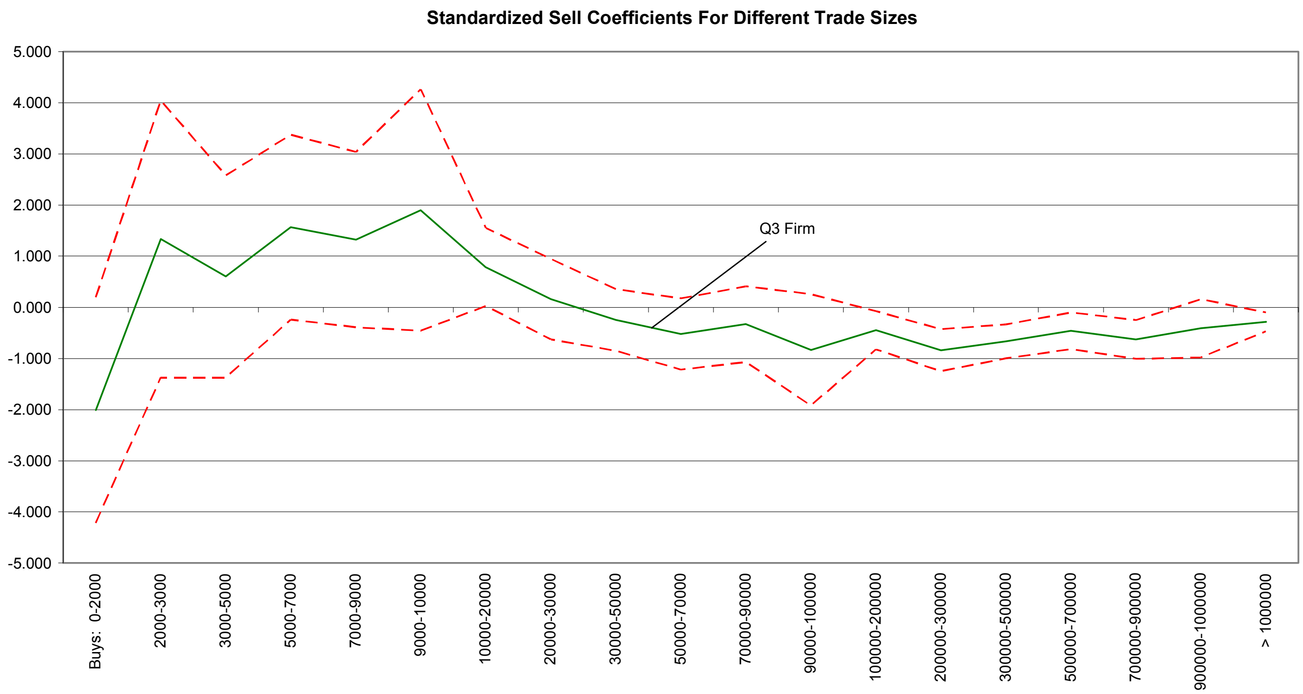


Figure 9

This figure plots the sell coefficients for each trade size bin, for the Q1, Q3 and Q5 firms in our sample. The coefficients are standardized by removing the within quintile cross-sectional mean of bin coefficients, and dividing by the cross-sectional standard deviation of bin coefficients.

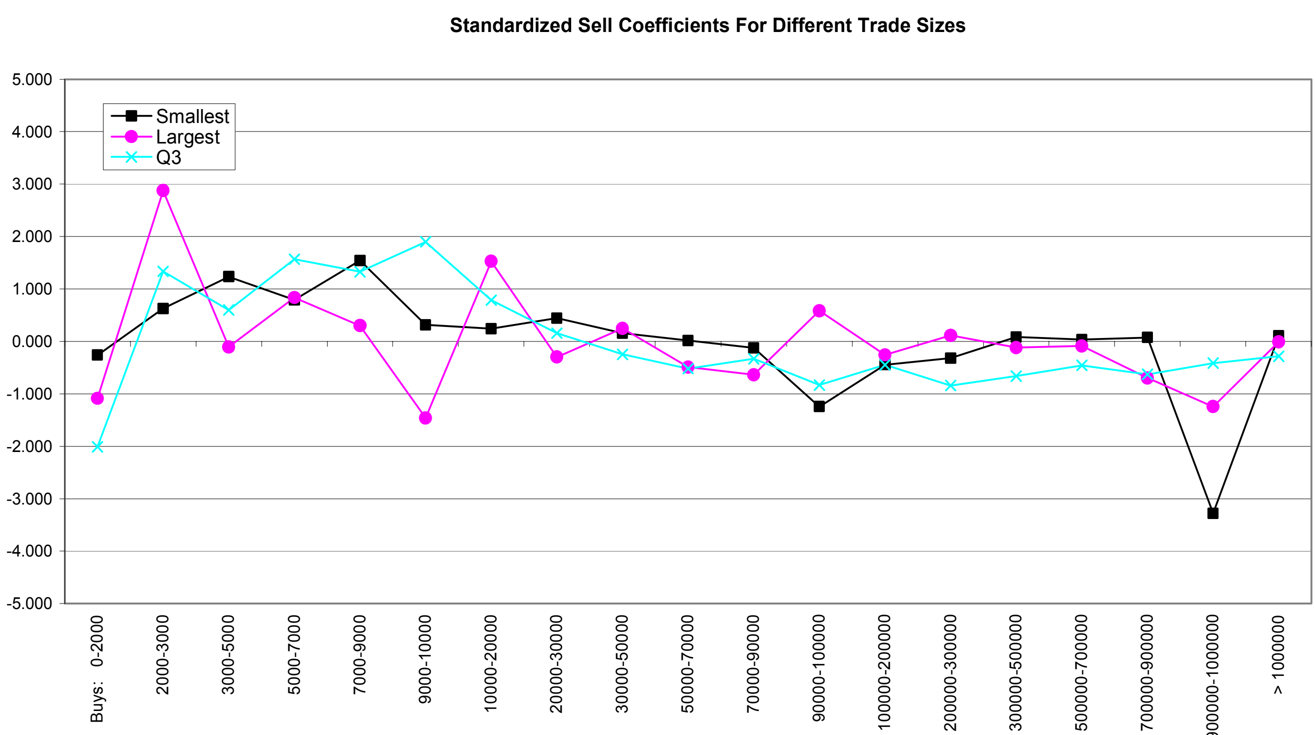


Figure 10

This figure plots the net flow coefficients estimated using the method of Nelson and Siegel [1987] for each trade size bin, for the Q1, Q3 and Q5 firms in our sample. The coefficients are standardized by removing the within quintile cross-sectional mean of bin coefficients, and dividing by the cross-sectional standard deviation of bin coefficients.

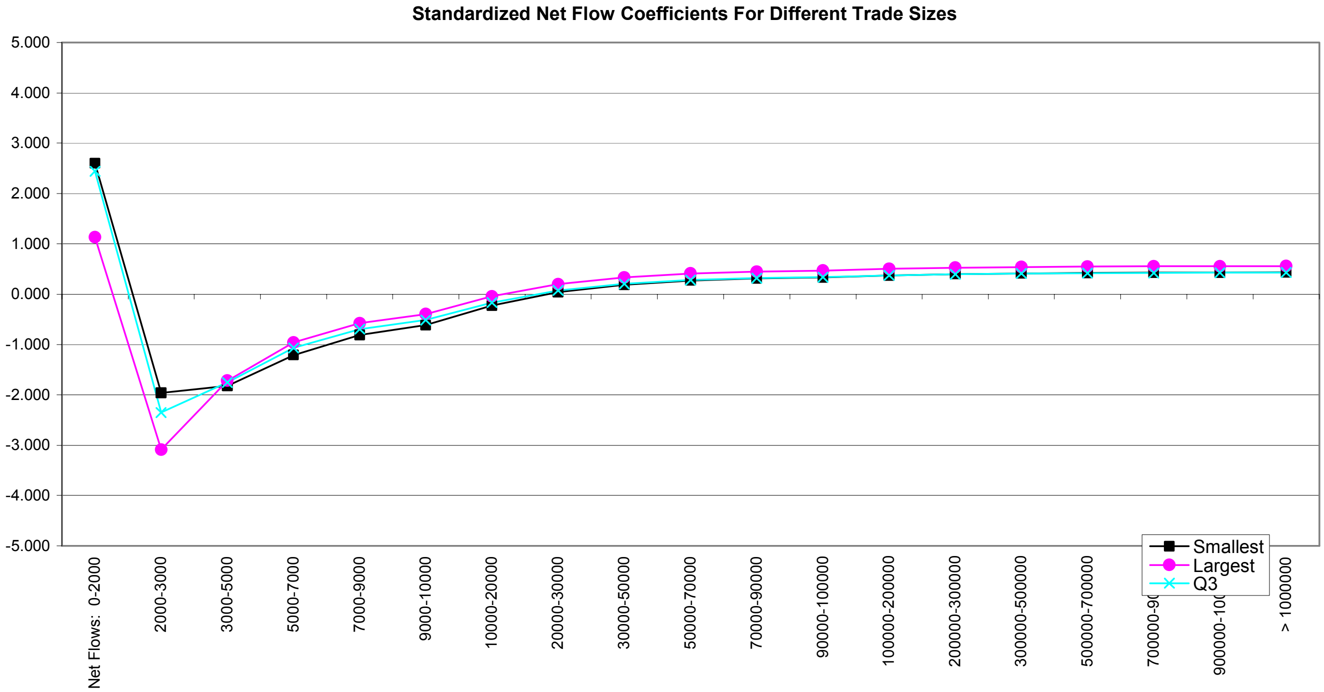


Figure 11

This figure plots the buy coefficients estimated using the method of Nelson and Siegel [1987] for each trade size bin, for the Q1, Q3 and Q5 firms in our sample. The coefficients are standardized by removing the within quintile cross-sectional mean of bin coefficients, and dividing by the cross-sectional standard deviation of bin coefficients.

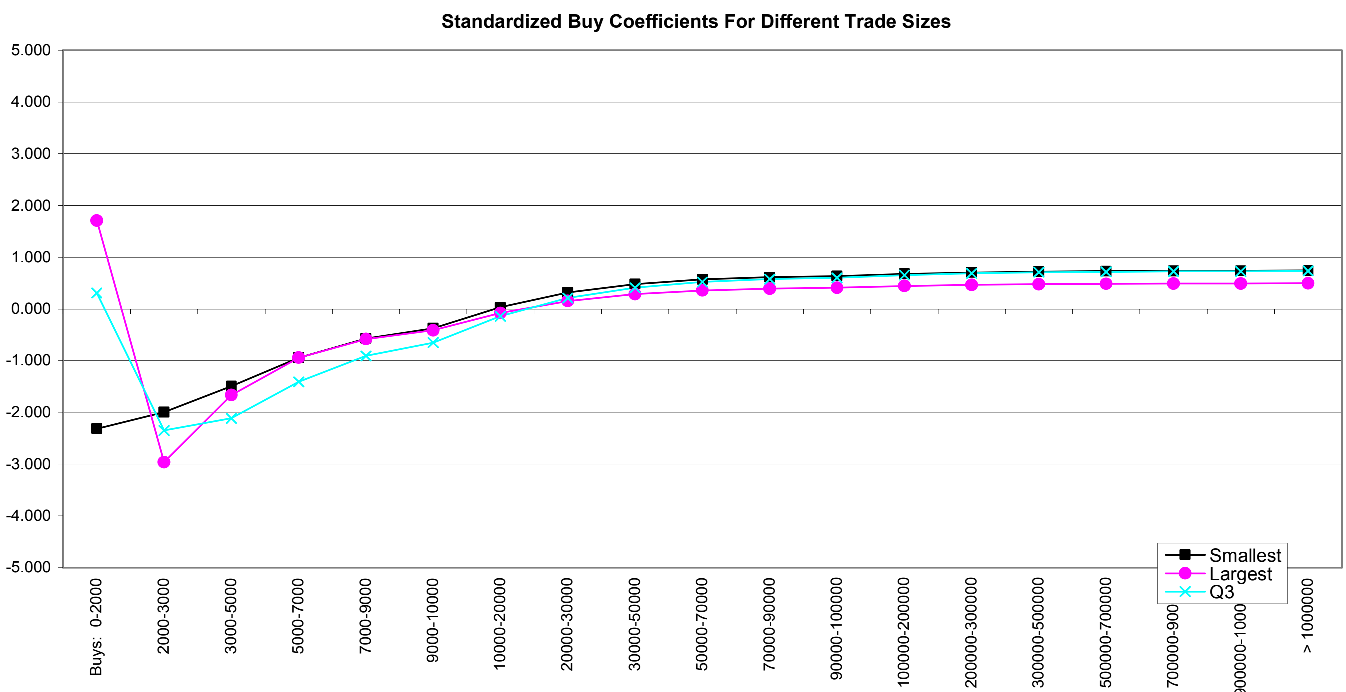


Figure 12

This figure plots the sell coefficients estimated using the method of Nelson and Siegel [1987] for each trade size bin, for the Q1, Q3 and Q5 firms in our sample. The coefficients are standardized by removing the within quintile cross-sectional mean of bin coefficients, and dividing by the cross-sectional standard deviation of bin coefficients.