CAUGHT ON TAPE: INSTITUTIONAL ORDER FLOW
AND STOCK RETURNS

John Y. Campbell
Tarun Ramadorai
Tuomo O. Vuolteenaho

Caught On Tape: Institutional Order Flow and Stock Returns
John Y. Campbell, Tarun Ramadorai, Tuomo O. Vuolteenaho
NBER Working Paper No. 11439
June 2005
JEL No. G1

ABSTRACT

Many questions about institutional trading can only be answered if one can track high-frequency changes in institutional ownership. In the US, however, institutions are only required to report their ownership quarterly in 13-F filings. We infer daily institutional trading behavior from the "tape", the Transactions and Quotes database of the New York Stock Exchange, using both a naive approach and a sophisticated method that best matches quarterly 13-F data. Increases in our measures of institutional flows negatively predict returns, particularly when institutions are selling. We interpret this as evidence that 13-F institutions compensate more patient investors for the service of providing liquidity. We also find that both very large and very small trades signal institutional activity, while medium size trades signal activity by the rest of the market.

John Y. Campbell
Department of Economics
Harvard University
Littauer Center 213
Cambridge, MA 02138
and NBER
john_campbell@harvard.edu

Tarun Ramadorai
Saïd Business School
University of Oxford
Park End Street
Oxford OX1 1HP
United Kingdom
tarun.ramadorai@sbs.ox.ac.uk

Tuomo O. Vuolteenaho
Arrowstreet Capital, LP
44 Brattle St., 5th floor
Cambridge, MA 02138
and NBER
tvuolteenaho@arrowstreetcapital.com

1. Introduction

This paper combines high-frequency information on equity transactions of different sizes with quarterly information on institutional equity holdings to infer daily movements in institutional equity ownership. Using these inferred daily changes in institutional ownership, or institutional flows, we provide new evidence on the short-term covariances between institutional flows and stock returns for a broad cross-section of stocks over the period 1995-2000.

Institutional equity holdings have interested finance economists ever since the efficient markets hypothesis was first formulated. One straightforward way to test the hypothesis is to inspect the portfolio returns of investors that are presumed to be sophisticated, such as mutual fund managers, to see if they earn more than a "fair" compensation for risk. Jensen (1968) pioneered this literature, finding little evidence to support the proposition that mutual fund managers earn abnormal returns. Many subsequent studies have examined the returns of mutual funds (e.g. Hendricks, Patel, and Zeckhauser (1993) and Carhart (1997)) or the returns on the portfolios that they report quarterly (e.g. Daniel, Grinblatt, Titman, and Wermers (1997), Wermers (2000)).

In recent years the literature on institutional holdings has moved in several new directions. First, other institutions besides mutual funds have been included in the investigation. Lakonishok, Shleifer, and Vishny (1992) examined the behavior of pension funds, Nofsinger and Sias (1999) looked at institutional equity owners as defined by Standard and Poors, and many recent papers have looked at all institutions that are required to make quarterly 13-F filings to the Securities and Exchange Commission. Second, the literature has examined the characteristics of stocks that institutional investors hold and not just their subsequent returns. Gompers and Metrick (2001) and Bennett, Sias, and Starks (2003), for example, run cross-sectional regressions of institutional ownership onto characteristics of individual stocks, documenting institutional preferences for large, liquid stocks and changes in those preferences over time. Third, there has been increased interest in the *changes* in institutional positions, their flows rather than their holdings. Quarterly institutional flows appear to be correlated with lagged quarterly stock returns (Grinblatt, Titman, and Wermers (1995), Badrinath and

Wahal (2002), Cai and Zheng (2004)), contemporaneous quarterly stock returns (Bennett, Sias, and Starks (2003)), and future quarterly stock returns (Wermers (1999) and Chen, Jegadeesh, and Wermers (2000) for mutual funds, and Bennett, Sias, and Starks (2003) for a broader set of institutions). Nofsinger and Sias (1999) find similar results at the annual frequency.

The literature on institutional flows is severely handicapped by the low frequency of the available data. While some countries, such as Finland (Grinblatt and Keloharju (2000a,b)) and Korea (Choe, Kho, and Stulz (1999)), do record institutional ownership almost continuously, in the United States institutional positions are reported only quarterly. This makes it hard to say whether institutions are reacting to stock price movements or causing price movements, as there is no resolution on the intra-quarter covariances of institutional flows and returns.

There has been some recent progress on measuring these intra-quarter covariances. Sias, Starks, and Titman (2001) point out that monthly return data can be combined with quarterly ownership data to make at least some inferences about monthly lead-lag relations between flows and returns. Boyer and Zheng (2004) apply this methodology to equity ownership data from the Flow of Funds accounts. The Sias-Starks-Titman approach ingeniously extracts additional information from quarterly data, but can only put bounds on monthly leads and lags, and has very little to say about lead-lag relations at higher frequencies than monthly.

A number of other papers have used proprietary datasets to measure high-frequency institutional behavior. Froot, O'Connell and Seasholes (2001), Froot and Ramadorai (2001), and Froot and Teo (2004) employ custodial data from State Street corporation, and find evidence of bidirectional positive Granger causality between weekly institutional flows and returns on equity portfolios in a variety of countries. Lee and Radhakrishna (2000) study the TORQ data set, a sample of trades with complete identification of market participants. Jones and Lipson (2003) and Kaniel, Saar and Titman (2004) employ Audit Trail data from the NYSE. The latter paper focuses on the behaviour of individual investors' trades, and

shows that individual investor purchases (sales) precede positive (negative) movements in stock returns. Jones and Lipson (2001) and Barber and Odean (2005) use weekly data from Plexus, a transactions cost measuring service for a subset of money managers. Griffin, Harris, and Topaloglu (2003) study the trades of NASDAQ brokerage houses that specialize in dealing with either individual or institutional investors, and find that institutions buy stocks that have recently risen, both at the daily frequency and the intra-daily frequency. These studies offer tantalizing glimpses of institutional behavior, but are limited in several respects. They are of course difficult to replicate, and their samples are typically restricted in their coverage of institutional investors, the cross-section of stocks they consider, the time span they investigate, or some combination thereof.

Another strand of the literature utilizes data on equity transactions available on the New York Stock Exchange Trade and Quotes (TAQ) database. Most transactions can be identified as buy orders or sell orders using the procedure of Lee and Ready (1991), which compares the transaction price to posted bid and ask quotes. A common procedure is to then separate trades by dollar size, identifying 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 the performance of several alternative cutoff rules in the TORQ data set. 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.[1]

This paper makes two substantial contributions to the literature. First, we evaluate the performance of the naive cutoff rule approach to measuring institutional flows, using the quarterly 13-F filings data as a benchmark. We show that we can greatly improve the explanatory power of this exercise by using the TAQ data in a more sophisticated fashion. Second, we construct both naive and sophisticated measures of daily institutional flows. We

---

[1]Several authors use a variant of the Lee and Radhakrishna approach. Ofek and Richardson (2003) use block trades as a measure of institutional participation in a stock. Hvidkjaer (2005) and Malmendier and Shanthikumar (2004) partition TAQ into small and large trades, identifying the former with individuals, and the latter with institutions.

find that flows respond to lagged daily stock returns positively, but predict future daily stock returns negatively, suggesting that institutions demand liquidity when they trade.

In order to perform our benchmarking exercise, we combine the TAQ database (the "tape") with the Spectrum database, which records the quarterly 13-F filings of large institutional investors. All institutions managing $100 million or more must report all long positions exceeding 10,000 shares or $200,000. Thus the Spectrum database measures the significant long holdings of large institutional investors; the complement of the Spectrum data includes short positions, extremely small institutional long positions, and the equity holdings of small institutions and individual investors. The nature of the Spectrum database is discussed in more detail in the Appendix. For simplicity in describing our results, we refer to 13-F filers as institutions, and the rest of the market as individuals.

The Lee-Radhakrishna approach performs poorly when benchmarked against the quarterly Spectrum data. For example, a cutoff rule that classifies all trades over $20,000 as institutional, and all trades under $2,500 as individual has a negative R^2 when used as a predictor of the change in institutional ownership reported in Spectrum. We find the best function mapping trade size to institutional behavior by regressing changes in institutional ownership on cumulative trades of different sizes. This more sophisticated approach generates a 12 percent R^2 in the same prediction exercise.

Why this enormous difference? The key insight is that a cutoff rule classifies each trade as institutional based on its own characteristics alone. This approach treats each trade as an independent event, and ignores the fact that each trade can help reveal the origins of other trades. The joint probability of the occurrence of trades of various sizes is the relevant statistic. This means that the coefficients on trade size bins in a regression predicting institutional ownership may be very different from the probabilities that trades of that size are institutional. In contrast with a cutoff rule, our analysis gives us a regression function that relates patterns in trading volume to changes in institutional ownership. We construct fitted values using the regression coefficients, and use these fitted values as a measure of high-frequency institutional equity flows.

Our approach reveals some important properties of institutional trading. First, institutions on average appear to 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. These results suggest that institutions use the liquidity provided by the specialist and possibly also provided by limit orders from individuals. Second, buying at the ask and selling at the bid is more likely to be indicative of institutional buying or selling if the trade size is either very small or very large. Trades that are either under \$2,000 or over \$30,000 in size reveal institutional activity, whereas intermediate size trades reveal individual activity. Finally, small trades are stronger indicators of institutional activity in stocks that already have a high level of institutional ownership.

Using both the naive cutoff approach and our more sophisticated method to construct daily institutional flows, we study the dynamics of institutional trading at the daily frequency in relation to daily stock returns. Daily institutional flows are highly persistent, consistent with the results of proprietary data analyses (e.g. Froot, O'Connell and Seasholes (2001) and Froot and Ramadorai (2001)) and with the persistence of total order imbalances reported by Chordia and Subrahmanyam (2004). Daily institutional flows are positively related to past daily stock returns, consistent with the quarterly evidence on momentum trading by institutions (Grinblatt, Titman, and Wermers (1995), Badrinath and Wahal (2002), Cai and Zheng (2004)).

We find that daily institutional flows predict *negative* and significant movements in subsequent market-adjusted stock returns, across the four largest size quintiles of stocks. This is consistent with the findings of Kaniel, Saar, and Titman (2004) that individual purchases predict positive returns at high frequencies, but at first glance it is surprising given the evidence for institutional outperformance at the quarterly frequency. We interpret our result as telling us that institutions demand liquidity when they trade. As in the model of Campbell, Grossman and Wang (1993), risk-averse liquidity providers accommodate high-frequency institutional demands for liquidity, but they require compensation for this service. The compensation takes the form of a price premium (discount) which generates a negative

(positive) stock return subsequent to institutional buying (selling).[2]

In a final exercise, we separately condition future returns on positive and negative flows. We find that institutional sales strongly predict positive returns, while institutional purchases only weakly predict negative returns. This asymmetry implies that the rest of the market is well compensated for accommodating institutional selling pressure, and less so for accommodating institutional buying pressure. The asymmetry is consistent with the presence of institutional constraints on short selling (see Hong and Stein (2003) and Nagel (2005)). An institutional investor can increase exposure to an underlying factor by purchasing a variety of alternative stocks. However, in the presence of short-sales constraints, it is impossible to reduce exposure to a factor other than by selling pre-existing holdings. The lack of substitutes for institutional sell transactions could result in their consuming greater liquidity than purchases. In addition, 13-F filers are not required to report short positions, so the asymmetry could reflect the existence of unmeasured institutional short sales that are correlated with measured institutional sales of long positions.

The organization of the paper is as follows. Section 2 introduces the TAQ, Spectrum and CRSP data used in the study, and discusses several interesting features of these data. Section 3 presents results from the benchmarking exercise, and introduces our method for predicting institutional ownership. In Section 4 we explain how the daily institutional flow measures are constructed, and relate these measures to daily stock returns. Section 5 concludes.

2. Data

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

---

[2]In a similar spirit, Kim (2000) and Chordia and Subrahmanyam (2004) find that net order imbalance negatively predicts returns, although Bennett and Sias (2001) find the opposite result.

current analysis, we focus on ordinary common shares of firms incorporated in the United States that traded on the NYSE and AMEX.[3] 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. 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. The number of matched firms in our data changes over time, as firms list or delist from the NYSE and AMEX, or move between NYSE and AMEX and other exchanges.

In the majority of our analysis, we present results separately 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 3329 firms. When sorted quarterly into size quintiles, this results in 735 firms in the largest quintile, and between 1125 and 1351 firms in the other four quintiles (these numbers include transitions of firms between quintiles), and 62,946 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 dynamically map each ticker symbol to a CRSP PERMNO. As tickers change over time, and are sometimes recycled or reassigned, this mapping also varies 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

---

[3]Ellis, Michaely and O'Hara (2000) show that the use of trade classification rules such as Lee and Ready (2000) in NASDAQ introduces biases in classifying large trades and trades initiated during high volume periods, especially for trades executed inside the spread.

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 of 'unclassifiable' trades.

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.[4] 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 most of our specifications, 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 then 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 a total of 102 such stock-quarters). We then winsorize each

---

[4]Finucane (2000) and Odders-White (2000) provide evidence that small trades, and trades in highly liquid stocks tend to be more frequently misclassified.

volume measure and the net order imbalance at the 1 and 99 percentile points. That is, we replace outliers with the 1st or 99th percentile points of the distribution.[5]

The differences in trading patterns across small and large stocks are summarized in Table I, which 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. Most of this difference manifests itself in the final years of our sample. The distribution of total volume is positively skewed within each quintile, so median volumes are somewhat lower. Nevertheless, median volumes also increase with market capitalization. This is consistent with the results of Lo and Wang (2000), who attribute the positive association between firm size and turnover to the propensity of active institutional investors to hold large stocks for reasons of liquidity and corporate control. The within-quintile annualized standard deviation of total volume (computed under the assumption that quarterly observations are iid) is fairly similar for stocks of all sizes, ranging from 30 percent to 36 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 –2.2 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.[6] This is consistent with the strong price performance of large stocks during most of this period.

Unclassifiable volume is on average about 16 percent of shares outstanding in our data set. 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

---

[5]We re-ran all our specifications with and without winsorization, and the results are qualitatively unchanged.

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

volume in the largest quintile. It is encouraging that the algorithm appears equally reliable among firms of different 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 1 summarizes the distribution of buy and sell volume across trade sizes. The figure reports three histograms: for the smallest, median, and the largest quintiles 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. The figure reveals 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.

2.3. Spectrum data

Our data on institutional equity ownership come from the Spectrum database, currently distributed by Thomson Financial. They have been 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. We exclude all stock-quarters for which either the level or change of Spectrum institutional ownership as a percentage of shares outstanding is greater than 100 percent (there are a total of 625 such stock-quarters). We then winsorize these data in the same manner as the TAQ data, at the 1 and 99 percentile points of the distribution of stock-quarters.

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 1.4 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.

These patterns may result in part from strong performance of institutionally held stocks, which moves these stocks into larger quintiles over time, but institutions have also been selling smaller stocks and buying larger stocks. This corresponds nicely with the trade intensity histograms in Figure 1, 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 anomalous in that these stocks have larger-size sales than buys despite their growth in institutional ownership.

3. Inferring High-Frequency Institutional Flows

3.1. Cutoff rules

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) (henceforth LR) evaluate alternative cutoff rules using the TORQ data set. 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. Unfortunately the TORQ data set 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.

We use an alternative benchmark to evaluate the method. We match the TAQ data at the trade sizes prescribed by different cutoff rules to the Spectrum data for a broad cross-

12

section of stocks, over our entire sample period. The cutoff model can be thought of as a restricted regression where buys (sells) in sizes above the upper cutoff get a coefficient of plus one (minus one), buys (sells) in sizes below the lower cutoff get a coefficient of minus one (plus one), and trades in intermediate sizes get a coefficient of zero.

We estimate this restricted regression in Table II, for a variety of cutoff values proposed by LR. In all cases we remove quarter-specific means, and allow free coefficients on both the lagged level and lagged change in institutional ownership on the right hand side of each regression, to soak up possible long-term mean reversion and short-term dynamics in institutional holdings. When the coefficient restrictions implied 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 is maximized at 4.6 percent, 5.6 percent, and 8.2 percent for the median, fourth and largest quintiles respectively.

In the second row under each cutoff value, the restrictions are relaxed, and the regression is allowed to freely estimate coefficients on the cutoff values proposed by LR. This causes the R^2 statistics of the regressions to increase substantially. The two smallest size quintiles' R^2 statistics are now maximized at 7.6 and 5.8 percent respectively, and those for the three larger quintiles are now all greater than 10 percent. This dramatic improvement suggests that the information available in the order flow data can be much better utilized. In particular, combining the information available in trade size bins is better than simply categorizing trades as institutional or individual without regard to the overall trading environment.

3.2. Why is a regression method better?

It is useful to understand the source of the improvement offered by a regression-based method over a cutoff rule approach. Consider the following example: Suppose all individuals trade in $10,000 amounts and trade in a perfectly correlated manner (either all sells, or all buys on a particular day); assume that every institution except for one trades in $10,000 amounts, in a manner that is perfectly positively correlated with all other institutions and perfectly nega-

tively correlated with individuals; finally one large institution trades in $100,000 amounts, in a manner that is perfectly correlated with all other institutions. In this case the probability that a $10,000 trade is institutional, based on its own characteristics is 50 percent, and the probability that a $100,000 trade is institutional is 100 percent. However, if we observe a $100,000 buy, then we can infer that all the $10,000 buys are institutional with probability 100 percent.

Translating this to the context of our regressions, this means that volume occurring in trade sizes of $100,000 should get a coefficient that is far greater than the unit coefficient that would be implied by a cutoff rule, because it reveals the direction of all the $10,000 institutional trades. This admittedly extreme example suggests that we can optimally use the information on the intra-quarter tape by combining various trade size bins in the way that best explains the quarterly changes in institutional ownership identified in Spectrum.

3.3. Basic regression method

As a preliminary step, we estimate extremely simple regressions that ignore the information in trade sizes, to see what we can learn about the data in the most restricted specification. 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 (all variables are expressed as percentages of the end-of-quarter t shares outstanding of stock i), we estimate:

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

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), short and long-run mean-reversion in the institutional share for particular stocks (the autoregressive coefficients ϕ and ρ), 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 imbalance $F_{it} = B_{it} - S_{it}$ and we estimate:

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

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, for the five quintiles of market capitalization. Across all size quintiles, buy volume gets a positive coefficient and sell volume gets a negative coefficient. 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 indicates that institutions are particularly likely to behave in this way when they are selling. 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 autoregressive coefficients are negative, and small but precisely estimated, telling us that there is statistically detectable mean-reversion in institutional ownership, at both short and long-run horizons.

The explanatory power of these regressions is U-shaped in market capitalization, above eight percent for the smallest firms, above 10 percent for the largest quintile, and around six percent for the median size firms. Note that simply allowing the regression to determine the appropriate sign and magnitude of the coefficients on unclassifiable volume and net order imbalance already generates performance improvements over the cutoff rule specifications in Table II, despite restricting the coefficients on every trade size bin to be the same.

3.4. The information in trade size

We generalize our specification to allow separate coefficients on net flows in each trade size bin:

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

where Z indexes trade size.

A concern about the specification (3.3) is that it requires 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 dollar bin sizes. 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:

$$\beta(Z) = b_0 + (b_1 + b_2)\left[1 - e^{-Z/\tau}\right]\frac{\tau}{Z} - b_2 e^{-Z/\tau}. \tag{3.4}$$

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 function $\beta(Z)$ approaches its limit b_0 as trade size Z increases. We also consider a variant of the Nelson-Siegel function that allows the effect of trade size Z to vary with an interaction variable ν_{it} :

$$\beta(Z, \nu_{it}) = b_{01} + b_{02}\nu_{it} + (b_{11} + b_{12}\nu_{it} + b_{21} + b_{22}\nu_{it})\left[1 - e^{-Z/\tau}\right]\frac{\tau}{Z} - (b_{21} + b_{22}\nu_{it})e^{-Z/\tau}. \tag{3.5}$$

To keep the model parsimonious, we do not allow the parameter τ to vary with ν_{it}.

Writing $g_1(Z) = \frac{\tau}{Z}(1 - e^{-Z/\tau})$ and $g_2(Z) = \frac{\tau}{Z}(1 - e^{-Z/\tau}) - e^{-Z/\tau}$, we can estimate the function using nonlinear least squares, searching over different values of τ, to select the

function that maximizes the R^2 statistic, resulting in:

$$\Delta Y_{it} = \alpha_{it} + \rho \Delta Y_{i,t-1} + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_{U\nu}\left(\nu_{i,t-1}U_{it}\right)$$
$$+ b_{01}\sum_Z F_{Zit} + b_{02}\sum_Z \nu_{i,t-1}F_{Zit} + b_{11}\sum_Z g_1(Z)F_{Zit}$$
$$+ b_{12}\sum_Z g_1(Z)\nu_{i,t-1}F_{Zit} + b_{21}\sum_Z g_2(Z)F_{Zit} + b_{22}\sum_Z g_2(Z)\nu_{i,t-1}F_{Zit} + \varepsilon_{it}. \quad (3.6)$$

Armed with the parameters of function (3.5), we can evaluate the function at different levels of ν_{it}, providing comparative statics on changes in institutional trading patterns with the interaction variable.

Robust standard errors in all cases are computed using the Rogers (1983, 1993) method. These standard errors are consistent in the presence of heteroskedasticity, cross-correlation and autocorrelation of up to one year in our quarterly specifications.[7]

Table IV estimates equation (3.6) separately for each quintile of market capitalization, replacing the intercept α with time dummies, and using the level of lagged institutional ownership $(Y_{i,t-1})$ as the interaction variable ν. The statistical significance of the estimated parameters is quite high, giving us some confidence in the precision of our estimates of the implied trade-size coefficients. Overall, the information in trade sizes adds considerable explanatory power to our regressions. Comparing the second panel in Table III with Table IV, the R^2 statistics increase from 8.3 percent to 12.3 percent in the smallest quintile, from 6.6 percent to 14.2 percent in the median quintile, and from 10.9 percent to 14.2 percent in the largest quintile. Of course, these R^2 statistics remain fairly modest, but it should not be surprising that institutional trading activity is hard to predict given the incentives that institutions have to conceal their activity, the considerable overlap between the trade sizes that may be used by wealthy individuals and by smaller institutions, and the increasing use of internalization and off-market matching of trades by institutional investors.

---

[7]We also computed heteroskedasticity and cross-contemporaneous correlation consistent standard errors using the nonparametric jackknife methodology of Shao and Wu (1989) and Shao (1989). The results are similar.

Figure 2 plots the trade-size coefficients implied by the estimates in Table IV, setting the lagged level of quarterly institutional ownership to its in-sample mean. 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. 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 and median quintiles of firms, but not for the largest firms in the sample. However, figure 3 reveals that when the lagged level of quarterly institutional ownership is set to one standard deviation above its quarterly mean, the coefficient on extremely small trades turns positive even for the largest stocks in the sample.

This is consistent with several possibilities. Institutions might break trades into extremely small sizes when they are "stealth trading" (trying to conceal their activity from the market), or institutions may use "scrum trades" to close out extremely small equity positions.[8] Another possibility is that institutions may use extremely small trades to test the liquidity of the market before trading in larger sizes. Figure 3 suggests that this behaviour may be exaggerated when institutions own a large percentage of the stock. It could also be the case 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 trades.

The parsimony of equation (3.6) is extremely useful, in that it permits a relatively straightforward investigation of changes in the functional form over time. This allows us to investigate the time stability of our regression coefficients, and to compare the out of sample

---

[8]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 counter to our result here.

forecasting power of our method to the R^2 statistics implied by the LR method. The last row of Table II shows the implied R^2 statistics generated by out of sample forecasts from the Nelson-Siegel specification. These out of sample R^2 statistics are computed by rolling through time, expanding the dataset in each step. We begin by estimating the model from the first quarter of 1993 until the final quarter of 1994, and construct an implied fitted value for the first quarter of 1995 using these estimated parameters. We then re-estimate the Nelson-Siegel function on the expanded dataset in each period, progressively forecasting one period ahead. Across all size quintiles of stocks, the resulting out of sample R^2 statistics are higher than either the restricted or unrestricted LR R^2 statistics.

4. High-Frequency Institutional Flows and Returns

4.1. Constructing daily institutional flows

We now analyze the relationship between our measures of daily institutional flows and stock returns. We can think of equation (3.6) as a daily function aggregated up to the quarterly frequency. Writing d for a daily time interval within a quarter t, and letting q_t represent the number of days within quarter t, the daily function is:

$$\Delta Y_{id} = \alpha_d + \rho_d \Delta Y_{i,t-1} + \phi_d Y_{i,t-1} + \beta_U U_{id} + \beta_{U\nu} Y_{i,t-1} U_{id}$$
$$+ b_{01} \sum_Z F_{Zid} + b_{02} \sum_Z Y_{i,t-1} F_{Zid} + b_{11} \sum_Z g_1(Z) F_{Zid}$$
$$+ b_{12} \sum_Z g_1(Z) Y_{i,t-1} F_{Zid} + b_{21} \sum_Z g_2(Z) F_{Zid} + b_{22} \sum_Z g_2(Z) Y_{i,t-1} F_{Zid} + \varepsilon_{id} \quad (4.1)$$

We make an assumption here in time-aggregating (4.1) up to the quarterly frequency to obtain equation (3.6) that the error in measured daily institutional ownership ε_{id} is uncorrelated at all daily leads and lags within a quarter with all of the right hand side variables in equation (4.1). This exogeneity assumption seems plausible, although it might be violated if for example institutions alter their daily trading strategies in response to the transactions volume in different size bins observed in recent days. The assumption guarantees that the

parameters of the daily function $b_{01}, b_{02}, b_{11}, b_{12}, b_{21}, b_{22}, \tau$ will be the same as those estimated at the quarterly frequency.

Having estimated equation (3.6), we can recover the parameters of equation (4.1) , and construct the fitted value $\mathrm{E}_d[\Delta Y_{id}]$ on each day d for each stock i. This is our measure of daily institutional flows. When we construct this fitted value, we are careful not to incorporate any purely quarterly parameters or variables (ρ, ϕ, α and ε) as we will be forced to make ad-hoc assumptions about the intra-quarter timing of events if we do so. We construct the fitted value in two different ways, using either the in-sample or out-of-sample parameters estimated in Table IV. Henceforth we term $\mathrm{E}_d[\Delta Y_{id}]$ the institutional 'flow' for stock i on day d, and denote it as f_{id}.

Table V presents descriptive statistics for daily stock returns, for our two daily flow measures, and for daily flows constructed using the LR method. To implement the LR method, we pick the cutoffs that yield the highest R^2 statistic from Table II for each quintile. For example, for the median size quintile of stocks, we subtract net order imbalance occurring in trade sizes below \$5,000 from net order imbalance occurring in trade sizes above \$100,000. The sample in all cases is restricted by the requirements of our out-of-sample estimation, beginning on the first trading day of January 1995, and ending in December 2000. All daily flow measures are winsorized at the 1 and 99 percentile points of the distribution across all stock-days in the sample, and all series are 'market-adjusted' by subtracting the daily cross-sectional mean.

There are several features of interest in Table V. First, the means of our flow measures tend to be smaller in absolute value than the means of the LR flows. Second, across all sets of flows, it appears that institutions have been buying into large-cap stocks, and selling out of small and mid-cap stocks. Interestingly, market-adjusted returns have also been negative in the three smallest size quintiles of stocks, and positive in the two largest size quintiles. Gompers and Metrick (2001) suggest that institutional buying has driven up the prices of large stocks, generating positive returns to these stocks. Third, median flows are greater than mean flows, with the exception of LR flows in the largest quintile, implying

that the distribution of flows is skewed to the left. This suggests that institutions trade more aggressively on days when they sell than on days when they buy stocks.

Fourth, the standard deviations of our two flow measures are similar in magnitude, and approximately half the size of the LR flow standard deviation. Fifth, the large standard deviation of returns, especially for small stocks, is unsurprising considering that these are close-to-close returns that incorporate the bid-ask-bounce, evinced by the large negative first daily autocorrelation of small stock returns. Sixth, flows are highly positively autocorrelated, echoing the finding from proprietary data (see Froot, O'Connell and Seasholes (2001)) that institutional flows appear highly persistent at daily and weekly frequencies. The persistence of flows is more pronounced for our flow measures than for the LR measure.

Seventh, the contemporaneous daily correlation between the flow and return measures is high and positive for the two largest quintiles of stocks for all three flow measures. This suggests that there are potentially interesting intra-day flow-return relationships that could profitably be investigated. Note that our method and the LR method yield differing signs for these contemporaneous correlations for the first two size quintiles of stocks. Finally, the contemporaneous correlations between the three definitions of flows indicate that the in and out-of-sample flows we construct are highly correlated with each other; that the LR flows are not very highly correlated with either set of our flows for the smallest stocks in the sample, and that the correlation between LR flows and both sets of our flows is increasing across size quintiles. This last observation is consistent with our finding in Table II that the explanatory power of LR flows for institutional ownership changes is increasing across size quintiles.

4.2. A vector autoregression for flows and returns

We now turn to a more systematic investigation of the relationship between daily institutional flows and returns. We are interested in the answers to several questions that have been posed in the theoretical and empirical literature pertaining to liquidity and trading. First, when daily institutional ownership changes, what is the impact on future returns? Is there an

asymmetric response to increases and decreases in institutional ownership? Second, what are the predictors of institutional flows at the daily frequency? Do past flows predict future flows? Do past returns predict future flows?

In order to answer these questions, we need a summary of the joint dynamics of flows and returns. The dynamics in daily data involve multiple lags, so a standard vector autoregression has a large number of free parameters. We impose parsimony on the system by modelling the explanatory variables as exponentially weighted moving averages (EWMA), writing for a stock i on a day d:

$$r_{id} = \alpha_r + \rho_f \left((1 - \lambda_f) \sum_{j=1}^{d-1} (\lambda_f)^j f_{i,d-j} \right) + \rho_r \left((1 - \lambda_r) \sum_{j=1}^{d-1} (\lambda_r)^j r_{i,d-j} \right) + \varepsilon_{id}^r, \qquad (4.2a)$$

$$f_{id} = \alpha_f + \phi_f \left((1 - \chi_f) \sum_{j=1}^{d-1} (\chi_f)^j f_{i,d-j} \right) + \phi_r \left((1 - \chi_r) \sum_{j=1}^{d-1} (\chi_r)^j r_{i,d-j} \right) + \varepsilon_{id}^f. \qquad (4.2b)$$

Here, f_{id} and r_{id} are 'market-adjusted' by subtracting the daily cross-sectional mean flow and return across all stocks, respectively. We estimate the parameters $\rho, \lambda, \chi, \phi$ by grid-searching over the parameter space of λ_r, λ_f (χ_r, χ_f) for returns (flows), estimating an OLS regression for each combination of these parameters. We then pick the $\rho_r^*, \lambda_r^*, \rho_f^*, \lambda_f^*$ ($\phi_r^*, \phi_f^*, \chi_r^*, \chi_f^*$) that generate the highest value of the R^2 statistic in the return (flow) equation.

In order to check whether daily returns respond differently to 'purchases' and 'sales', we use the same λ_r^*, λ_f^* estimated in (4.2a),and estimate separate coefficients ρ_f^b and ρ_f^s on positive and negative weighted flows. The superscripts b and s denote "buys" and "sells", but what matters here is the sign of the moving average of flows, rather than the sign of the flow on any given day:

$$r_{id} = \alpha_r + \rho_f^b \left| \left((1 - \lambda_f) \sum_{j=1}^{d-1} (\lambda_f)^j f_{i,d-j} \right) \right|_{>0} + \rho_f^s \left| \left((1 - \lambda_f) \sum_{j=1}^{d-1} (\lambda_f)^j f_{i,d-j} \right) \right|_{<0}$$

$$+ \rho_r \left((1 - \lambda_r) \sum_{j=1}^{d-1} (\lambda_r)^j r_{i,d-j} \right) + \varepsilon_{id}^r. \quad (4.3)$$

If returns respond identically to institutional buying and to institutional selling, we should find no evidence that ρ_f^b and ρ_f^s are different.

We re-estimated these specifications skipping one day to avoid any potential contamination of our results by bid-ask bounce, and found that our results were essentially unchanged. In all cases we construct robust standard errors for our daily specifications using the Rogers (1983, 1993) method, and verify that the results are similar when we use a contemporaneous cross-correlation consistent jackknife estimator. We estimated equations (4.2) and (4.3) for all three measures of flow, and found that the results were surprisingly quite similar. In what follows, we present results for our out-of-sample flows.

The top and bottom panels of Table VI present estimates of equations (4.2) and (4.3). The upper half of the top panel estimates equation (4.2b), revealing several interesting features. First, it is clear that the flows are highly persistent. χ_f is estimated to be 0.9 for every size quintile of stocks, translating to a half-life of 6.6 days. The ϕ_f coefficients are all estimated to be around 0.7, and highly statistically significant. Flow persistence appears to be virtually identical across size quintiles of stocks.

In the market microstructure literature, persistent flows are generally thought to characterize the trading behaviour of informed investors (Kyle (1985)). However, our measures represent changes in the *aggregate* ownership of institutional investors. Given this, the persistence we find could also emanate from daily lead-lag effects across the trades of different institutional investors (see Hong, Kubik and Stein (2004, 2005)).

Second, flows respond positively to lagged returns across all size quintiles. The coefficient is small, but precisely estimated. In the largest size quintile of stocks, for example, a

220 basis point increase in returns (approximately one daily standard deviation) over the previous two days would generate an approximately 1.2 basis point increase in the flow (approximately one-fifth of a standard deviation). This has variously been called trend-chasing or return-chasing behaviour (see Froot et. al (2001) and Froot and Ramadorai (2001) for two recent examples). However another valid non-behavioral explanation is that institutional investors have rationally revised their expectations upward in response to good cash-flow news, resulting in portfolio rebalancing.

The bottom half of the top panel of Table VI presents estimates of equation (4.2a). Stock returns in all but the median quintile of stocks are negatively related to the EWMA of past returns. This is robust to lagging the independent variables an additional day to avoid bid-ask bounce, and has been described as the 'weekly reversal' effect in the literature (see Subrahmanyam (2005) and references therein). Turning to the coefficients on lagged flows, we find that estimates of λ_f vary dramatically across size quintiles, but our results are not an artefact of λ_f variation; when we impose the restriction that $\lambda_f = 0.9$ for every size quintile, our results are qualitatively unchanged.[9] The coefficients on lagged flows are negative and statistically significant across each of the four largest size quintiles of stocks. The magnitude of the coefficient for the median quintile implies that a one basis point increase in flow relative to the cross-sectional mean daily flow causes a 'market-adjusted' decrease of around 50 basis points in expected stock returns. This short-term relationship has the opposite sign from the positive long-term relationship typically found in the literature on quarterly institutional behavior.

Our measure represents the daily net flows of institutional investors, which should be negatively correlated with measures of individual investor purchases. Thus our result mirrors that of Kaniel, Saar and Titman (2004) who find that individual investor flows positively forecast returns at the weekly frequency. One explanation that is consistent with the findings in this paper and those of Kaniel et. al. can be found in Campbell, Grossman and Wang (1993), who model the interaction between groups of investors that have different

---

[9]We find that the likelihood function is very flat with respect to variation in λ_f.

propensities to take and provide liquidity. If risk-averse individual investors (and other non-13-F filers) act like market makers and accommodate high-frequency institutional demands for immediacy, some compensation will be required for providing this service. Here, the compensation shows up as a negative (positive) stock return subsequent to institutional buying (selling).

The second panel in Table VI presents estimates of equation (4.3). In each of the four larger size quintiles of stocks, both ρ_f^s and ρ_f^b are positive, indicating that institutional purchases and sales both predict higher returns. However ρ_f^s is much larger than ρ_f^b (in most cases more than twice the size), and in the two largest size quintiles of stocks, only ρ_f^s is statistically significant. These results tell us that in the larger size quintiles, institutional sales move prices against institutions but institutional purchases do not. Results are quite different in the smallest quintile, where institutional sales predict lower returns. This could be evidence that institutions provide liquidity for individual purchases in small stocks, for which individual investors have a noted preference (see Barber and Odean (2000)).

The asymmetry in larger stocks is strongly statistically significant, and also consistent with the results of Kaniel, Saar, and Titman (2004) who find that buying by individual investors predicts positive returns, but that individual selling does not predict negative returns. The asymmetry could result from an inability or reluctance of many institutional portfolio managers to use short sales. When institutions wish to increase exposure to an underlying factor, they can substitute from one stock to others if the price of their preferred purchase runs up too much. As a result, institutional buy transactions are not likely to consume a great deal of liquidity. However, in the absence of short sales, if institutions wish to reduce exposure to the same underlying factor, the only way to do so is to sell the specific stocks purchased earlier. This suggests a reason why institutional stock sales will consume more liquidity than institutional purchases. Of course, some institutions do sell stocks short and these positions are not reported on 13-F forms. Our results could also reflect unmeasured short selling pressure that is correlated with measured institutional sales.

5. Conclusion

Using both naive and sophisticated methods to infer high-frequency institutional ownership, we provide new evidence on the daily covariances between institutional portfolio flows and stock returns for a broad cross-section of stocks over the 1995 to 2000 period.

We find that our measures of institutional flows predict future daily stock returns negatively. Most of this negative predictability is the result of institutional sales predicting high subsequent returns. We explain the negative return predictability as an implicit payment by liquidity-demanding institutions to liquidity providers, as in the model of Campbell, Grossman and Wang (1993). The sign asymmetry suggests that institutional sales demand more liquidity than institutional purchases. We also find, in accordance with much of the literature employing proprietary datasets, that our measures of institutional flows are highly persistent, and follow movements in daily returns. Institutions buy stocks that have recently done well, and sell those that have done poorly.

Our second main contribution is to evaluate the performance of cutoff-rule based techniques to infer institutional ownership from publicly available transactions data. We find that these techniques perform poorly when benchmarked against the quarterly 13-F filings required of institutional investors by the SEC. We develop a new method in this paper that represents a marked improvement over the standard cutoff-rule approach, even when we use our method out of sample. As a by-product of this benchmarking exercise, we discover new results pertaining to the trading behaviour of institutional investors. 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, test the liquidity of the market, or to round small positions up or down. All these patterns are reversed for sell volume, and are remarkably consistent across firm sizes.

There has been a great deal of recent interest in the returns and trading patterns of institutional investors. This paper attempts to answer some of the open questions in this

26

area. In future research we plan to use our method to investigate institutional trading behavior around a variety of corporate actions such as earnings announcements, stock splits, and dividend initiations and omissions.

6. Appendices

6.1. 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

28

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.

6.2. 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. Institutions must report regardless of whether they are regulated by the SEC, and foreign institutions must report if they use any means of United States interstate commerce. Holdings are reported quarterly on the SEC's form 13-F, 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 13-F 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 13-F. On the 13-F, 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.

The Spectrum data on institutional equity positions are incomplete in two respects. First, some institutions receive "confidential treatment". Each quarter the SEC's Division of Investment Management reviews requests from money managers anxious to keep some or all of their holdings from being publicly disclosed. Confidential treatment can be granted on either a partial or complete basis. The SEC then withholds that quarter's confidential information for one year before it is made public. According to journalistic reports, the SEC

generally grants confidential treatment exemptions for proprietary investment methodologies that would be in jeopardy if holdings were disclosed on a regular basis. When the confidential treatment exemption expires, these data are not subsequently backfilled by Spectrum. Second, institutions are not required to report short positions. Given that the majority of institutional investors (pension funds, mutual funds, insurance companies) have investment mandates preventing short sales, this will affect our inferences to the extent that hedge funds or proprietary traders hold short positions.

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 13-F 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.

6.3. Appendix 3: Robustness of Bin Definition

The flexibility of the Nelson-Siegel functional form allows us to check whether our specification can be improved by alternative definitions of trade size bins. We currently define our

bins in terms of the dollar size of a trade. This dollar based bin classification is motivated by the insight that we can use the wealth constraint experienced by individuals to try to separate the trading behaviour of institutions from that of individuals. In other words, individual investors generally either cannot trade large dollar trade sizes because they simply don't have the money, or dislike making large dollar trades because such trades would result in extremely concentrated and/or leveraged positions relative to their wealth.

Another possible constraint we could use to separate individuals from institutions is the liquidity constraint, i.e. institutions generally do not like to trade illiquid securities for a variety of reasons (such as the desire to window dress their portfolios). This, especially for active institutional traders, indicates a preference for more liquid trade sizes in which it is easier to increase or decrease holdings.[10] This in turn suggests that we redefine our bins each quarter in terms of percentiles of total trading volume that fall within each bin. Yet another approach is to specify bins in terms of multiples of average quoted depth, as a measure of the 'normal' or 'most liquid' trade size in a stock. We used a straightforward way to check whether the liquidity constraint can help us better identify institutional ownership - we interacted our dollar size bins with measures of liquidity - total daily volume, and average quoted depth.

When we replace the lagged institutional ownership interaction with these liquidity interactions, we find that they do contribute incremental explanatory power over the function (3.4). However, our specification in Table IV is robust to incorporating these additional liquidity interactions. These measures of liquidity contribute no incremental explanatory power over the lagged institutional ownership interaction. This gives us confidence that our final specification is robust to movements in daily liquidity.

---

[10]Thanks to Soeren Hvikdjaer for first bringing this issue to our attention.

References

Badrinath, S.G. and Wahal, Sunil, 2002, Momentum trading by institutions, *Journal of Finance* 57, 2449–2478.

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

Bennett, James and Richard Sias, 2001, Can money flows predict stock returns?, *Financial Analysts Journal* 57 (6), 64–78.

Bennett, James, Richard Sias, and Laura T. Starks, 2003, Greener pastures and the impact of dynamic institutional preferences, *Review of Financial Studies* 16, 1199–1234.

Boyer, Brian and Lu Zheng, 2004, Who moves the market? A study of stock prices and sector cashflows, unpublished paper, University of Michigan.

Cai, Fang and Lu Zheng, 2004, Institutional trading and stock returns, *Finance Research Letters* 1, 178–189.

Campbell, John Y., Sanford J. Grossman and Jiang Wang, 1993, Trading volume and serial correlation in stock returns, *Quarterly Journal of Economics* 108, 905–939.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

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

Chen, Joseph, Harrison Hong and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* 66, 171–205.

Chen, Hsiu-Lang, Narasimhan Jegadeesh, and Russ 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.

Choe, Hyuk, Bong-Chan Kho, and René M. Stulz, 1999, Do foreign investors destabilize stock markets? The Korean experience in 1997, *Journal of Financial Economics* 54, 227–264.

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

Cohen, Randy, Paul Gompers, and Tuomo O. Vuolteenaho, 2002, Who underreacts to cashflow news? Evidence from trading between individuals and institutions, *Journal of Financial Economics* 66, 409–462.

Daniel, Kent, D., Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* 52, 1035–1058.

Ellis, K., Roni Michaely., and Maureen O'Hara, 2000, The accuracy of trade classification rules: evidence from Nasdaq, *Journal of Financial and Quantitative Analysis* 35, 529–551.

Finucane, Thomas J., 2000, A direct test of methods for inferring trade direction from intra-day data, *Journal of Financial and Quantitative Analysis* 35, 553–576.

Froot, Kenneth A., Paul O'Connell, and Mark Seasholes, 2001, The portfolio flows of international investors, *Journal of Financial Economics*, 59, 151–193.

Froot, Kenneth A. and Tarun Ramadorai, 2001, The information content of international portfolio flows, NBER working paper 8472.

Froot, Kenneth A. and Melvyn Teo, 2004, Equity style returns and institutional investor flows, NBER working paper 10355

Gompers, Paul A., and Andrew Metrick, 2001, Institutional investors and equity prices, *Quarterly Journal of Economics* 116, 229–260.

Greene, Jason, 1995, The impact of limit order executions on trading costs in NYSE stocks: an empirical examination, unpublished paper, Indiana University.

Griffin, John M., Jeffrey H. Harris, and Selim Topaloglu, 2003, The dynamics of institutional and individual trading, *Journal of Finance* 58, 2285–2320.

Grinblatt, Mark and M. 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.

Grinblatt, Mark and M. Keloharju, 2000b, What makes investors trade?, *Journal of Finance* 56, 589–616.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers, 1995, Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior, *American Economic Review* 85, 1088–1105.

Hendricks, Daryl, Jayendu Patel and Richard Zeckhauser, 1993, Hot hands in mutual funds: short-run persistence of relative performance, 1974-1988, *Journal of Finance* 48, 93–130.

Hong, Harrison and Jeremy C. Stein, 2003, Differences of opinion, short-sales constraints, and market crashes, *Review of Financial Studies* 16, 487–525.

Hong, Harrison, Jeffrey Kubik and Jeremy C. Stein, 2004, Social interaction and stock-market participation, *Journal of Finance* 59, 137-163.

Hong, Harrison, Jeffrey Kubik and Jeremy C. Stein, 2005, Thy neighbor's portfolio: word-of-mouth effects in the holdings and trades of money managers, forthcoming *Journal of Finance*.

Hvidkjaer, Soeren, 2005, A trade-based analysis of momentum, forthcoming *Review of Financial Studies*.

Jensen, Michael C., 1968, The performance of mutual funds in the period 1945-1964, *Journal of Finance* 23, 389–416.

Jones, Charles M. and Marc L. Lipson, 2001, Sixteenths: direct evidence on institutional execution costs, *Journal of Financial Economics* 59, 253–278.

Jones, Charles M. and Marc L. Lipson, 2003, Are retail orders different?, unpublished paper, Columbia University and University of Georgia.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2004, Individual investor sentiment and stock returns, unpublished paper, Duke University, New York University, and University of Texas.

Kim, Jung-Wook, 2000, An analysis of the impact of trades on stock returns using money flow data, unpublished paper, Harvard University.

Kovtunenko, Boris and Nathan Sosner, 2003, Sources of institutional performance, unpublished paper, Harvard University.

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

Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, 1992, The impact of institutional trading on stock prices, *Journal of Financial Economics* 32, 23–43.

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

Lee, Charles M. C. and Balkrishna Radhakrishna, 2000, Inferring investor behavior: evidence from TORQ data, *Journal of Financial Markets* 3, 83–111.

Lo, Andrew W. and Jiang Wang, 2000, Trading volume: definitions, data analysis, and implications of portfolio theory, *Review of Financial Studies* 13, 257–300.

Malmendier, Ulrike and Devin Shanthikumar, 2005, Are small investors naive about incentives?, unpublished paper, Stanford University.

Nagel, Stefan, 2005, Short sales, institutional investors, and the cross-section of stock returns, *Journal of Financial Economics*, forthcoming.

Nelson, Charles R. and Andrew F. Siegel, 1987, Parsimonious modeling of yield curves, *Journal of Business* 60, 473–489.

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

Odders-White, Elizabeth R., 2000, On the occurrence and consequences of inaccurate trade classification, *Journal of Financial Markets* 3, 259–286.

Ofek, Eli and Matthew Richardson, 2003, DotCom mania: the rise and fall of internet stock prices, *Journal of Finance* 58, 1113–1138.

Rogers, W. H., 1983, Analyzing complex survey data, unpublished paper, Rand Corporation, Santa Monica, CA.

Rogers, W.H., 1993, Regression standard errors in clustered samples, Stata Technical Bulletin Reprints STB-13:STB-18, 88–94.

Shao, J., 1989, The efficiency and consistency of approximations to the jackknife variance estimators, *Journal of the American Statistical Association* 84, 114–119.

Shao, J. and C.F.J. Wu, 1989, A general theory for jackknife variance estimation, *Annals of Statistics* 17, 1176–1197.

Sias, Richard, 2004, Institutional herding, *Review of Financial Studies* 17, 165–206.

Sias, Richard, Laura T. Starks, and Sheridan Titman, 2001, The price impact of institutional trading, unpublished paper, Washington State University and University of Texas at Austin.

Subrahmanyam, Avanidhar, 2005, Distinguishing between rationales for short-horizon predictability in stock returns, *Financial Review* 40, 11–35.

Wermers, Russ, 1999, Mutual fund herding and the impact on stock prices, *Journal of Finance* 54, 581–622.

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

This table presents means, medians and standard deviations for the TAQ and Spectrum variables in our specifications. Both TAQ and Spectrum data were filtered to remove outliers (details in the appendix), and winsorized at the 1 and 99 percentile points. 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 for firm size quintiles, where firms are sorted quarterly by market capitalization (size), followed by those for all firms.

	Small	**Q2**	**Q3**	**Q4**	**Large**	**All**
Mean						
TAQ Total Buys	21.42	28.04	34.09	39.72	38.55	32.37
TAQ Total Sells	23.76	29.16	33.58	36.89	34.10	31.50
TAQ Unclassifiable	9.87	13.49	16.25	19.12	19.58	15.66
TAQ Total Volume	55.19	70.76	84.00	95.85	92.33	79.64
TAQ Net Imbalance	-2.19	-1.09	0.54	2.86	4.48	0.92
Spectrum Change	-1.36	0.27	1.69	1.46	0.80	0.57
Median						
TAQ Total Buys	13.69	18.66	24.67	31.06	30.41	23.70
TAQ Total Sells	15.77	20.47	25.37	29.59	27.28	23.78
TAQ Unclassifiable	5.70	8.66	11.39	14.87	15.72	11.51
TAQ Total Volume	36.37	48.95	62.54	76.28	74.01	60.13
TAQ Net Imbalance	-1.23	-0.63	0.10	1.59	3.05	0.53
Spectrum Change	-0.03	0.41	1.64	1.34	0.99	0.43
Standard Deviation						
TAQ Total Buys	12.18	14.52	15.23	15.63	14.23	14.81
TAQ Total Sells	12.54	13.86	14.11	14.03	12.34	13.59
TAQ Unclassifiable	6.17	7.37	7.72	7.75	6.93	7.44
TAQ Total Volume	29.95	34.73	35.96	36.39	32.69	34.84
TAQ Net Imbalance	5.08	5.26	5.38	5.29	4.38	5.24
Spectrum Change	7.53	9.36	9.79	9.40	7.79	8.84

Table II
Evaluating the Lee-Radhakrishna Method

This table presents adjusted R-squared statistics for regressions that explain the change in Spectrum institutional ownership with institutional trading estimated using cutoff rules from Lee and Radhakrishna (2000). Flows above the upper cutoff in each case are considered institutional, while flows below the lower cutoff are regarded as originating from individuals. In the rows labeled 'restricted coefficients,' the coefficient on flows above the upper cutoff is constrained to be +1 while the coefficient on flows below the lower cutoff is constrained to be -1. In the 'free coefficients' specifications, the coefficients on upper and lower cutoffs are estimated in the regression. All specifications contain the lagged level and change in institutional ownership on the right hand side, and incorporate quarter-specific time dummy variables. All TAQ and Spectrum variables are expressed in percentages of the shares outstanding of the firm. The second to last row shows the in-sample adjusted R-squared statistics using the method presented in this paper (we term this the 'CRV method') from Table IV. The final row shows the one period ahead out of sample adjusted R-squared of the CRV method estimated from a regression updated with one additional period each step, beginning with the first eight calendar quarters in the dataset.

Lower Cutoff; Upper Cutoff Adjusted R-Squared →	Small	Q2	Q3	Q4	Large
Lower = 2,000; Upper = 5,000					
Restricted Coefficients	-0.121	-0.120	-0.057	-0.002	0.038
Free Coefficients	0.069	0.043	0.067	0.081	0.101
Lower = 3,000; Upper = 10,000					
Restricted Coefficients	-0.080	-0.090	-0.030	0.011	0.046
Free Coefficients	0.074	0.050	0.082	0.093	0.112
Lower = 3,000; Upper = 20,000					
Restricted Coefficients	-0.050	-0.059	-0.003	0.025	0.054
Free Coefficients	0.076	0.055	0.089	0.097	0.115
Lower = 3,000; Upper = 50,000					
Restricted Coefficients	-0.021	-0.030	0.026	0.044	0.068
Free Coefficients	0.074	0.058	0.095	0.100	0.118
Lower = 5,000; Upper = 100,000					
Restricted Coefficients	-0.024	-0.016	0.046	0.056	0.082
Free Coefficients	0.068	0.057	0.103	0.104	0.122
CRV In Sample	0.123	0.100	0.142	0.133	0.142
CRV Out of Sample	0.108	0.101	0.130	0.130	0.131

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

This table presents results from a regression of the change in Spectrum institutional ownership on flows constructed from TAQ, estimated separately for stocks sorted into market capitalization quintiles. The dependent variable in all specifications is the change in Spectrum institutional ownership. The first panel below presents the independent variables in rows: the lagged level of Spectrum institutional ownership (*LS*), the lagged change in institutional ownership (*Δ(LS)*), the total unclassifiable volume in TAQ (*TAQ UC*), 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 TAQ and Spectrum variables are expressed in percentages of the shares outstanding of the firm. All specifications incorporate quarter-specific time dummy variables. Robust t-statistics computed using the Rogers (1983, 1993) method are reported in italics below the coefficients.

	Small	Q2	Q3	Q4	Large
LS	-0.043	-0.025	-0.018	-0.016	-0.022
	-7.802	*-7.189*	*-6.062*	*-4.358*	*-6.056*
Δ(LS)	-0.049	-0.020	-0.030	-0.073	-0.159
	-1.882	*-0.986*	*-1.667*	*-3.570*	*-5.868*
TAQ UC	-0.075	0.028	0.012	0.017	0.014
	-1.279	*0.486*	*0.213*	*0.337*	*0.283*
TAQ Total Buys	0.154	0.205	0.353	0.473	0.557
	4.342	*4.791*	*9.394*	*12.020*	*14.276*
TAQ Total Sells	-0.215	-0.293	-0.451	-0.559	-0.661
	-6.236	*-6.878*	*-10.669*	*-13.261*	*-15.432*
Adjusted R-Squared	0.084	0.049	0.069	0.083	0.113
N	12427	12526	12529	12632	12832
N(Firms)	1125	1351	1305	1162	735
Time Dummies?	Yes	Yes	Yes	Yes	Yes

	Small	Q2	Q3	Q4	Large
LS	-0.043	-0.026	-0.021	-0.018	-0.025
	-7.810	*-7.562*	*-6.979*	*-5.265*	*-6.810*
Δ(LS)	-0.048	-0.019	-0.028	-0.071	-0.157
	-1.841	*-0.939*	*-1.599*	*-3.498*	*-5.826*
TAQ UC	-0.164	-0.116	-0.146	-0.116	-0.141
	-5.054	*-3.208*	*-4.758*	*-3.926*	*-5.551*
TAQ Net Flows	0.188	0.246	0.387	0.498	0.575
	5.996	*6.172*	*10.877*	*13.196*	*15.172*
Adjusted R-Squared	0.083	0.046	0.066	0.081	0.109
N	12427	12526	12529	12632	12832
N(Firms)	1125	1351	1305	1162	735
Time Dummies?	Yes	Yes	Yes	Yes	Yes

Table IV
Estimates of Nelson-Siegel Function Coefficients

This table presents nonlinear least squares estimates of the Nelson-Siegel (1987) function that relates the change in quarterly 13-F institutional ownership from Spectrum to exogenous variables, TAQ flows and an interaction with the lagged institutional ownership fraction. The independent variables are: the lagged level of Spectrum institutional ownership (LS), the lagged change in Spectrum institutional ownership ($\Delta(LS)$), the total unclassifiable volume in TAQ ($TAQ\ UC$), $TAQ\ UC$ interacted with LS, bin specific TAQ flows, and bin specific TAQ flows interacted with LS. All TAQ and Spectrum variables are expressed in percentages of the shares outstanding of the firm. The coefficients on flows in various bins (indexed by Z, the midpoint of the range of dollar trade sizes captured in the bin) can be recovered from the coefficients below. The function:

$$\beta(Z, LS) = (b_{01} + b_{02}LS) + (b_{11} + b_{12}LS + b_{21} + b_{22}LS)[1 - e^{-Z/\tau}]\frac{\tau}{Z} - (b_{21} + b_{22}LS)e^{-Z/\tau}$$

All specifications incorporate quarter-specific time dummy variables. Robust t-statistics computed using the Rogers (1983, 1993) method are reported in italics below the coefficients. We refer to our method as the 'CRV method', accordingly we report the 'CRV Adjusted R-Squared.'

	Small	Q2	Q3	Q4	Large
Control Variables					
LS	0.066	0.223	0.346	0.378	0.262
	2.163	*5.480*	*5.326*	*4.599*	*2.923*
$\Delta(LS)$	-1.085	-0.801	-0.782	-0.758	-0.560
	-8.508	*-7.056*	*-6.853*	*-6.035*	*-4.280*
TAQ UC	-0.009	0.005	0.012	0.014	0.005
	-1.561	*1.499*	*3.072*	*3.687*	*1.069*
(TAQ UC)*(LS)	-0.051	-0.028	-0.041	-0.085	-0.169
	-1.960	*-1.413*	*-2.545*	*-4.310*	*-6.310*
Nelson-Siegel Coefficients					
b_{01}	0.157	0.261	0.551	0.591	0.848
	2.579	*3.086*	*5.707*	*4.740*	*4.373*
b_{02}	0.360	0.330	0.029	-0.068	-0.366
	1.691	*2.214*	*0.195*	*-0.359*	*-1.277*
b_{11}	4.180	7.389	26.235	3.893	-4.985
	1.261	*1.776*	*2.205*	*1.266*	*-0.490*
b_{12}	41.195	25.226	-10.526	-21.858	-9.764
	1.370	*1.740*	*-0.415*	*-4.053*	*-0.634*
b_{21}	-5.710	-12.484	-39.865	-7.400	5.620
	-1.340	*-2.122*	*-2.635*	*-1.325*	*0.359*
b_{22}	-56.431	-42.580	-0.153	26.100	6.671
	-1.567	*-2.242*	*-0.005*	*2.783*	*0.286*
τ	498.302	984.704	989.769	5030.480	5031.451
CRV Adjusted R-Squared	0.123	0.100	0.142	0.133	0.142
N	12427	12526	12529	12632	12832
N(Firms)	1125	1351	1305	1162	735
Time Dummies?	Yes	Yes	Yes	Yes	Yes

Table V
Summary Statistics for Daily Flows and Returns

This table presents means, medians, standard deviations, the first daily autocorrelation, and the contemporaneous daily correlation between flows and daily stock returns from CRSP, and correlations between different flow measures for the three types of daily flows we construct using the TAQ data. These are the 'Lee-Radhakrishna Flows', estimated using the best restricted cutoff rule specification for each size quintile chosen from Table II, flows constructed using the coefficients we estimate in Table IV ('CRV In Sample Flows'), and flows constructed using out-of-sample estimated coefficients from our method ('CRV Out of Sample Flows'). All flow and return measures are 'market-adjusted' by subtracting the daily cross-sectional mean across all stocks. All flow measures are winsorized at the 1 and 99 percentile points across all stock-days, and are in basis points of daily shares outstanding as reported in CRSP. Daily returns are expressed in basis points. The columns report these summary statistics for firm size quintiles, where firms are sorted quarterly by market capitalization (size).

	Small	Q2	Q3	Q4	Large
Mean					
Lee-Radhakrishna Flows	-0.415	-0.703	-0.441	0.330	0.937
CRV In Sample Flows	-0.011	-0.237	-0.311	0.107	0.461
CRV Out of Sample Flows	-0.120	-0.147	-0.294	0.111	0.426
Returns	-4.512	-2.599	-0.987	1.260	3.133
Median					
Lee-Radhakrishna Flows	-0.258	-0.301	-0.288	-0.116	0.325
CRV In Sample Flows	0.417	0.374	0.351	0.468	0.496
CRV Out of Sample Flows	0.160	0.206	0.190	0.307	0.349
Returns	-19.159	-16.161	-13.068	-10.381	-7.035
Standard Deviation					
Lee-Radhakrishna Flows	6.993	10.565	13.114	13.135	9.631
CRV In Sample Flows	3.051	5.753	7.855	7.820	5.978
CRV Out of Sample Flows	3.345	5.890	7.595	7.628	5.904
Returns	591.953	352.249	290.776	247.139	220.597
First Daily Autocorrelation					
Lee-Radhakrishna Flows	0.072	0.060	0.090	0.138	0.212
CRV In Sample Flows	0.193	0.167	0.168	0.195	0.241
CRV Out of Sample Flows	0.182	0.171	0.177	0.202	0.252
Returns	-0.162	-0.008	0.018	-0.005	-0.002
Corr(Flows(t),Returns(t))					
Lee-Radhakrishna Flows	-0.062	0.084	0.162	0.231	0.338
CRV In Sample Flows	0.067	-0.026	0.070	0.206	0.319
CRV Out of Sample Flows	0.021	-0.062	0.028	0.172	0.293
Corr(CRV In Sample(t), CRV Out of Sample(t))	0.893	0.955	0.978	0.979	0.982
Corr(CRV In Sample(t), Lee-Radhakrishna(t))	0.488	0.674	0.803	0.870	0.903
Corr(CRV Out of Sample(t), Lee-Radhakrishna (t))	0.474	0.635	0.776	0.879	0.912

Table VI
Vector Autoregression of Daily Flows and Returns

This table presents estimates of a bivariate VAR system (with an EWMA coefficient restriction) of daily institutional investor flows constructed using out-of-sample coefficients from the method in this paper ('CRV Out of Sample Flows'), and daily stock returns. Flows and returns are cross-sectionally demeaned each day to 'market adjust'. The top panel estimates the equation:

$$f_{i,d} = \alpha_f + \phi_f \left(1 - \chi_f\right) \sum_{j=1}^{d-1} \chi_f^j f_{i,d-j} + \phi_r \left(1 - \chi_r\right) \sum_{j=1}^{d-1} \chi_r^j r_{i,d-j} + \varepsilon_{i,d}^f$$

$$r_{i,d} = \alpha_r + \rho_f \left(1 - \lambda_f\right) \sum_{j=1}^{d-1} \lambda_f^j f_{i,d-j} + \rho_r \left(1 - \lambda_r\right) \sum_{j=1}^{d-1} \lambda_r^j r_{i,d-j} + \varepsilon_{i,d}^r$$

While the bottom panel looks at lagged 'buy' transactions separately from lagged 'sell' transactions in the return equation:

$$r_{i,d} = \alpha_r + \rho_f^b \left[\left(1 - \lambda_f\right) \sum_{j=1}^{d-1} \lambda_f^j f_{i,t-j} \right]_{>0} + \rho_f^s \left[\left(1 - \lambda_f\right) \sum_{j=1}^{d-1} \lambda_f^j f_{i,d-j} \right]_{<0} + \rho_r \left(1 - \lambda_r\right) \sum_{j=1}^{d-1} \lambda_r^j r_{i,d-j} + u_{i,d}^r$$

Columns contain estimates of firm size-quintile-specific coefficients. Rogers (1983, 1993) robust t-statistics are presented in *italics* below coefficients.

	Small	Q2	Q3	Q4	Large
	$f_{i,t}$	$f_{i,t}$	$f_{i,t}$	$f_{i,t}$	$f_{i,t}$
χ_f	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**
ϕ_f	0.695	0.653	0.656	0.678	0.699
	38.521	*57.057*	*52.577*	*64.587*	*50.510*
χ_r	**0.75**	**0.90**	**0.75**	**0.50**	**0.50**
ϕ_r	0.0006	0.0044	0.0051	0.0044	0.0056
	7.5970	*7.3113*	*11.1277*	*16.8946*	*14.2939*
R^2	0.079	0.069	0.075	0.085	0.113
	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$
λ_f	**0.98**	**0.00**	**0.00**	**0.90**	**0.90**
ρ_f	1.975	-0.976	-0.469	-0.605	-0.805
	1.736	*-6.918*	*-6.159*	*-2.579*	*-3.060*
λ_r	**0.50**	**0.75**	**0.00**	**0.75**	**0.75**
ρ_r	-0.322	-0.075	0.018	-0.035	-0.050
	-19.058	*-4.531*	*3.247*	*-3.674*	*-3.468*
R^2	0.0284	0.0011	0.0005	0.0003	0.0005

|	Small	Q2	Q3	Q4	Large
	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$
λ_f	**0.98**	**0.00**	**0.00**	**0.90**	**0.90**
ρ_f^b	-4.319	1.662	0.653	0.293	0.578
	-1.289	*5.579*	*4.357*	*1.020*	*1.813*
ρ_f^s	-3.624	2.531	1.167	1.330	2.611
	-2.570	*11.143*	*9.998*	*2.819*	*4.640*
λ_r	**0.50**	**0.75**	**0.00**	**0.75**	**0.75**
ρ_r	-0.322	-0.077	0.017	-0.036	-0.051
	-19.049	*-4.696*	*3.124*	*-3.862*	*-3.594*
R^2	0.0284	0.0018	0.0008	0.0003	0.0007
N	549,223	578,013	591,216	618,954	666,433

Figure 1

This figure 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).

Histogram of Trade Intensities - Q1 Firms
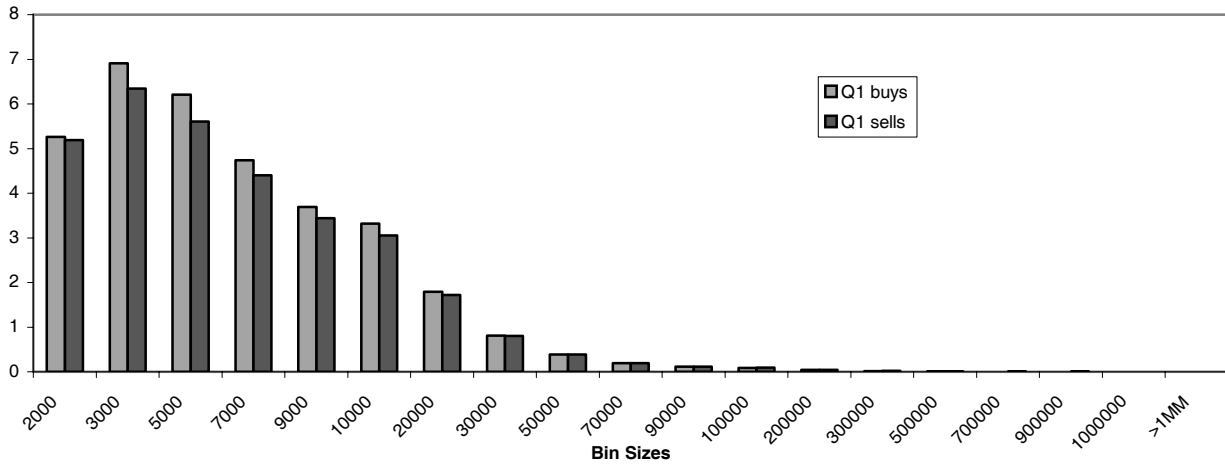
Histogram of Trade Intensities - Q3 Firms
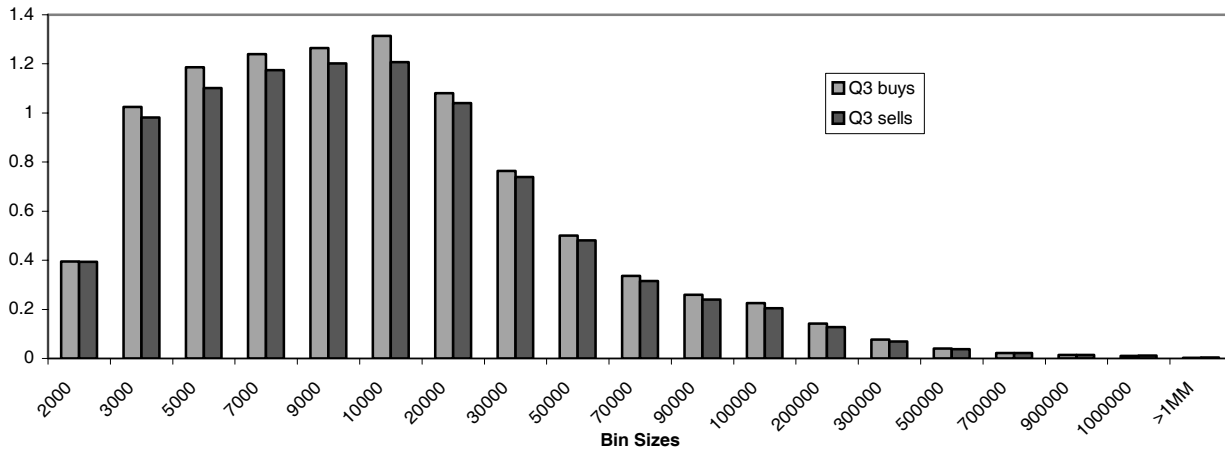
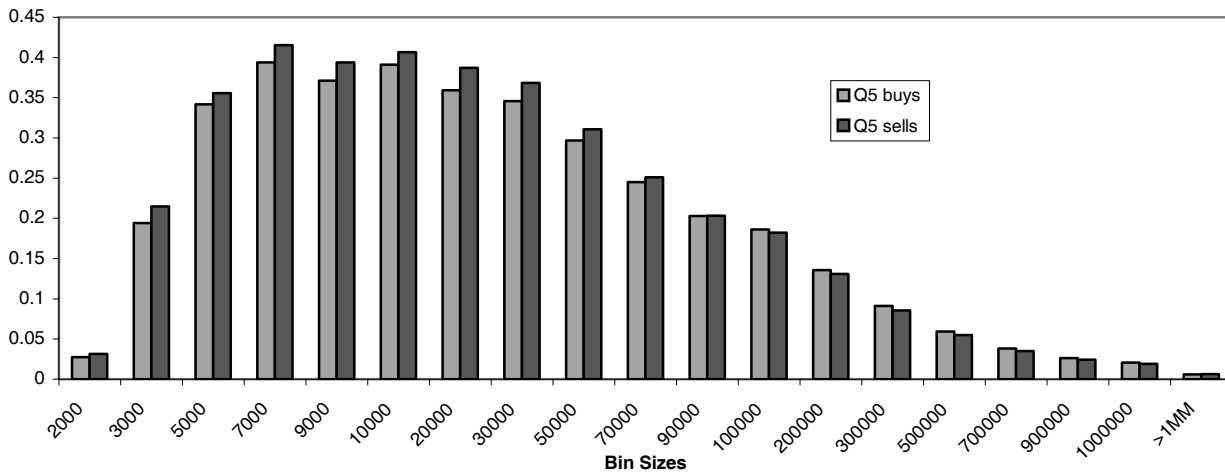Histogram of Trade Intensities - Q5 Firms

Figure 2

This figure plots the net flow coefficients estimated using the results in Table III 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.

**Standardized Net Flow Coefficients For Different Trade Sizes
At Mean Level of Lagged Quarterly Institutional Ownership**

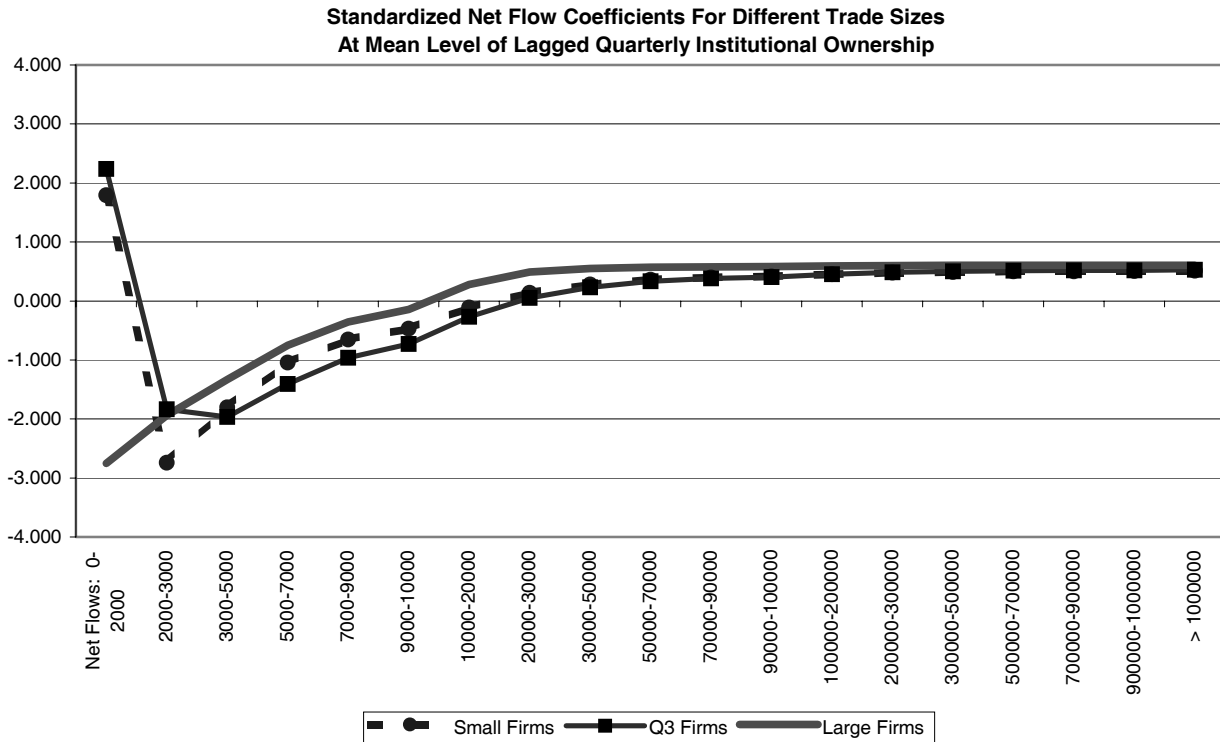Legend: ● Small Firms   ■ Q3 Firms   — Large Firms

Figure 3

This figure plots the net flow coefficients estimated using the results in Table III for each trade size bin, for the Q5 firms in our sample, setting the value of lagged institutional ownership to its quarterly mean and to one standard deviation above its quarterly mean.

**Raw Net Flow Coefficients For Different Trade Sizes
at Mean and Mean + 1 S.D. of LS: Largest Quintile**

Legend: — Mean   ■ Mean + 1 S.D.