Caught On Tape: Institutional Trading, Stock Returns, and Earnings Announcements

John Y. Campbell, Tarun Ramadorai, and Allie Schwartz[1]

July 2007

[1]Campbell: Department of Economics, Littauer Center 213, Harvard University, Cambridge MA 02138, USA, and NBER. Tel 617-496-6448, email john_campbell@harvard.edu. Ramadorai: Said Business School, University of Oxford, Park End Street, Oxford OX1 1HP, UK. Tel 44-1865-278816, email tarun.ramadorai@sbs.ox.ac.uk. Schwartz: Department of Economics, Littauer Center, Harvard University, Cambridge MA 02138, USA. Email schwart2@fas.harvard.edu. This paper replaces "Caught on Tape: Institutional Order Flow and Stock Returns", written by Campbell, Ramadorai, and Tuomo Vuolteenaho, and circulated as NBER Working Paper No. 11439. We are grateful to Tuomo Vuolteenaho for his many intellectual contributions to this paper. We thank Peter Hawthorne, Jakub Jurek, and Sung Seo for excellent research assistance; Boris Kovtunenko and Nathan Sosner for their assistance with the Spectrum dataset; Pablo Casas-Arce, Soeren Hvidkjaer, Pete Kyle, David Myatt, Narayan Naik, Venkatesh Panchapagesan, Kevin Sheppard, Joshua White, Pradeep Yadav and seminar participants at conferences for useful discussions. This material is based upon work supported by the National Science Foundation under Grant No. 0214061 to Campbell, and by Morgan Stanley and Co. under its Microstructure Research Program.

Caught On Tape: Institutional Trading, Stock Returns, and Earnings Announcements

Abstract

Many questions about institutional trading can only be answered if one can track high-frequency changes in institutional ownership. In the U.S., 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 a sophisticated method that best matches quarterly 13-F data. We find that daily institutional trades are highly persistent and respond positively to recent daily returns but negatively to longer-term past daily returns. Institutional trades, particularly sells, appear to generate short-term losses—possibly reflecting institutional demand for liquidity—but longer-term profits. One source of these profits is that institutions anticipate both earnings surprises and post-earnings-announcement drift. These results are different from those obtained using a standard size cutoff rule for institutional trades.

1. Introduction

How do institutional investors trade in equity markets? Do they hold stocks that deliver high average returns? Do they arbitrage apparent equity market inefficiencies such as post-earnings announcement drift, the tendency for stocks to continue to move in the same direction after an earnings announcement? More generally, are institutions a stabilizing or destabilizing influence on stock prices? These questions have been the focus of a large empirical literature.

In the United States, institutional investors are required to report their equity positions quarterly in 13-F filings to the Securities and Exchange Commission. These quarterly data show that changes in institutional equity holdings are positively serially correlated, positively correlated with lagged stock returns, and positively correlated with future stock returns. That is, institutions trade persistently, they buy recent winners and sell recent losers as momentum traders would do, and their trades are profitable on average. Contemporaneously, changes in institutional equity holdings are positively correlated with stock returns and earnings growth, but it is hard to know how to interpret these correlations because institutional trading can both drive stock returns and react to stock returns within the quarter, and can predict or follow earnings announcements.

To get a clearer picture of institutional trading patterns, one would like to be able to measure changes in institutional ownership as they occur. An obvious way to do this is to infer changing institutional ownership from equity transactions of different sizes. Several authors have done this assuming that large trades, above a fixed cutoff size, are institutional. In this paper we estimate a function mapping trades of different sizes into implied changes in institutional ownership, and find that the optimal function fits quarterly changes in institutional ownership much better than the cutoff rules that have been used in previous research.

Our method reveals some important properties of institutional trading. 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 consume liquidity. 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.

We use our method to infer daily institutional flows, and provide new evidence on the relationship between daily institutional trading, daily stock returns, and earnings surprises for a broad cross-section of US stocks over the period 1995-2000. We have five main findings. First, daily institutional trading is highly persistent, consistent with the quarterly evidence. Second, daily institutional trading reacts positively to recent daily returns, but negatively to longer-term past daily returns. This suggests that institutions are high-frequency momentum traders but contrarian investors at somewhat lower frequencies, a result not found in quarterly data. Third, daily institutional trading predicts near-term daily returns negatively, and longer-term daily returns positively. The latter result is consistent with the quarterly evidence that institutions trade profitably, but the former result suggests that institutions demand liquidity when they trade, moving stock prices in a manner that reverses the next day. Fourth, there is an asymmetry in this reversal. Next-day returns are significantly positive for institutional sales but not significantly negative for institutional purchases, suggesting that institutions demand more liquidity when they sell than when they buy. Fifth, institutional trading anticipates both earnings surprises and post-earnings-announcement drift (PEAD). That is, institutions buy stocks in advance of positive earnings surprises and sell them in advance of negative surprises; and the stocks they buy tend to experience positive PEAD while the stocks they sell tend to experience negative PEAD.

We compare these results with those that would be obtained using the standard cutoff rule approach. Basic findings such as trading persistence and the positive effect of very recent returns on institutional trades are common to both approaches. Many other findings, however, such as the negative effect of longer-term past returns on institutional trades, the tendency for short-term reversal and the longer-term profitability of institutional trading are much stronger and more consistent across all categories of stocks when we use our method for inferring institutional order flow. Finally, the predictive ability of institutional order flow for the earnings surprise and PEAD does not survive when cutoff rule based flows are used in place of flows created using our method.

The organization of the paper is as follows. The remainder of the introduction relates our paper to previous literature. Section 2 describes the TAQ, Spectrum and CRSP data used in the study. Section 3 introduces our method for predicting institutional ownership and compares it with a standard cutoff rule. Section 4 uses our method to construct daily institutional flows, and estimates a vector autoregression to describe the short- and medium-run dynamics of these flows. Section 5 asks how daily institutional flows anticipate and respond to earnings announcements. Section 6 concludes.

1.1. Related literature

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 positively correlated with lagged institutional flows (Sias (2004)), lagged quarterly stock returns (Grinblatt, Titman, and Wermers (1995), Badrinath and Wahal (2002), Cai and Zheng (2004)), contemporaneous quarterly stock returns (Grinblatt, Titman, and Wermers (1995), Wermers (1999, 2000), Nofsinger and Sias (1999), and Bennett, Sias, and Starks (2003)), and future quarterly stock returns (Daniel, Grinblatt, Titman, and Wermers (1997), 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 interpretation of these results is actively debated. Some authors, notably Lakonishok, Shleifer, and Vishny (1992), suggest that institutional investors follow simple price-momentum strategies that push stock prices away from fundamental values. This is disputed by others, such as Cohen, Gompers, and Vuolteenaho (2002), who find that institutions are not simply following price-momentum strategies; rather, they sell shares to individuals when a stock price increases in the absence of any news about underlying cash flows.

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 (2000 a,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 (2006) 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 (2007), and Froot and Teo (2007) employ custodial data from State Street corporation, and find evidence of flow persistence and 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 behavior 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 (2007) 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. Odean (1998, 1999) and Barber and Odean (2000, 2001, 2007) use data from a discount brokerage, and show that individual investors appear to over-trade and underperform. 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. The proprietary data may also be subject to selection bias if institutions self-select into transactions-cost measuring services or custodial pools.

There have been many previous attempts to use publicly available data from the New York Stock Exchange to measure high-frequency institutional equity trading. Kraus and Stoll (1972), Holthausen, Leftwich, and Mayers (1987), Madhavan and Cheng (1997), Ofek and Richardson (2003) and many others have used block trades as a measure of institutional participation in a stock. Much of this work seeks to estimate the price impact of block trades; Holthausen, Leftwich, and Mayers (1987) find that block sales temporarily depress stock prices, consistent with our fourth major finding.[2] Of course, block trades account for only a modest fraction of trading volume, and in recent years the Trade and Quotes (TAQ) database has allowed researchers to look at smaller equity trades.

Most transactions in the TAQ database can be identified as buys or sells 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 (lower) cutoff size as institutional (individual), with an intermediate buffer zone of medium sized trades that are not classified. Lee and Radhakrishna (2000) evaluate the performance of several alternative cutoff rules in the TORQ data set. They find, for example, that a $20,000 cutoff most effectively classifies institutional trades in small stocks. Hvidkjaer (2006) and Malmendier and Shanthikumar (2004) have followed a similar approach; they partition TAQ into small, medium and large trades using the Lee-Radhakrishna cutoff values. They acknowledge the Lee-Radhakrishna identification of small trades with individuals, and large trades with institutions, but prefer the monikers 'small traders' and 'large traders'.

Many of the same issues arise in the literature on post-earnings announcement drift (PEAD). This phenomenon has been well documented for a long time, so one would expect that sophisticated investors, including institutions, trade to take advantage of it. Indeed, Bartov et. al. (2000) find that PEAD is strongest in firms with low institutional share-

[2] Chan and Lakonishok (1993) and Keim and Madhavan (1995) also find asymmetric price impact of institutional purchases and sales using proprietary data.

holdings. Cohen, Gompers and Vuolteenaho (2002) find that institutions sell shares to individuals when a stock price increases in the absence of any news about underlying cash flows. Their measure of cash-flow news is obtained from a vector-autoregressive decomposition of unexpected stock returns. Ke and Ramalingegowda (2005) show that actively trading institutional investors move their stockholdings in the same direction as unexpected earnings and earn abnormal returns in subsequent quarters. While these results suggest that institutional investors act to take advantage of PEAD, their precision is somewhat limited by the low frequency of the data. A quarterly data frequency makes it hard to say whether institutions are reacting to stock price movements or causing price movements in the days surrounding earnings announcements.

Hirshleifer, Myers, Myers and Teoh (2004) use proprietary weekly data from a discount brokerage service and provide evidence that individual investors are significant net buyers after both negative and positive unexpected earnings. They do not find evidence that individuals' net trades have predictive power for future abnormal stock returns. Although this is useful evidence, it is hard to replicate, and subject to the selection bias inherent in the use of proprietary data from a single discount brokerage firm.

Lee (1992), Bhattacharya (2001), and Shanthikumar (2004) all use variants of the Lee-Radhakrishna method to study institutional trading around earnings announcements. Shanthikumar (2004) for example, finds that the imbalance between small purchases and small sales is unresponsive to the direction of unexpected earnings in the first month after an earnings announcement. In contrast, the imbalance between large purchases and large sales has the same sign as unexpected earnings. Shanthikumar interprets this finding as consistent with large traders' informational superiority, and with attempts by such traders to take advantage of PEAD. However, she finds that large trader order flow in the three days surrounding the earnings announcement forecasts the drift with a negative coefficient.

In this paper we evaluate the performance of the Lee-Radhakrishna cut-off rule using 13-F filings data as a benchmark. 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. The Spectrum database measures the significant long holdings of large institutional investors (we refer to these as "institutions"); the complement of the Spectrum data includes short positions, extremely small institutional long positions, and the equity holdings of small institutions and individual investors (for simplicity, we refer to this complement as "individuals"). We find that 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 has a negative adjusted R^2 when used as a predictor of the change in institutional ownership reported in Spectrum. In response to this finding we develop a superior method for identifying institutional order flow in section 3, apply it to the dynamics of daily institutional trading in section 4, and apply it to earnings announcements in section 5.

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 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,

[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.

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 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

[4] Finucane (2000) and Odders-White (2000) provide evidence that small trades, and trades in highly liquid

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 the net order imbalances in each size bin at the 1 and 99 percentile points. That is, we replace the outliers in each trade size bin with the 1st or 99th percentile points of the (pooled) distribution across all stock quarters.[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 55 percent of shares outstanding in the smallest quintile to 92 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

stocks tend to be more frequently misclassified.

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

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 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

[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.

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. Again, 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 pooled 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 has caused these stocks to move 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 Institutional Trading Behavior

3.1. Cutoff rules

In the market microstructure literature, institutional trading behavior has generally been identified using a cutoff rule. In particular, trades above an upper cutoff size are classified as originating from institutional investors, while those below a lower cutoff are classified as initiated by individual investors. 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-section of stocks, over our entire sample period. The cutoff model can be thought of as a restricted regression where the left-hand side variable is the quarterly change in Spectrum-reported institutional ownership, and on the right-hand side of the regression, buys (sells) in sizes above the cutoff get a coefficient of plus one (minus one) and trades in smaller 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 adjusted R^2 statistic in most cases is negative. In fact, the adjusted R^2 statistic given the restrictions on the flows above and below the cutoffs is never positive for the two smallest size quintiles, and maximized at 3.8 percent, 5.2 percent and 7.9 percent for the median, fourth and largest quintiles respectively.

In the second block of results in the table, the restrictions are relaxed, and the regression is allowed to freely estimate coefficients on the cutoff values proposed by LR. This causes the adjusted R^2 statistics of the regressions to increase substantially. The two smallest size quintiles' adjusted R^2 statistics are now at 6.7 and 5.3 percent respectively, and those for the three larger quintiles now range between 8.4 and 11.1 percent. This dramatic improvement suggests that the information available in the order flow data can be much better utilized. We explore the reasons for this improvement in the next sub-section.

3.2. Why is a regression method better?

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 negatively 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. This also implies that the regression coefficients cannot be interpreted as the probabilities of trades being institutional or individual.

Institutions have incentives to avoid detection by intermediaries (Kyle (1985)) and by methods such as ours, and they utilize order-splitting techniques to disguise their trades (Bertsimas and Lo (1998)). It is, therefore, an empirical question whether institutions are successful in avoiding such detection. We now turn to our empirical specifications.

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:

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

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:

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

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

Table III reports estimates of equation (3.1) in the top panel, and equation (3.2) in the bottom panel, 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 initiate trades, 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 six percent for the median size firms, and above 10 percent for the largest 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 now generalize our specification to allow separate coefficients on net flows in each trade size bin:

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

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:

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

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 variation of the Nelson-Siegel function which not only varies with trade size Z, but also with an interaction variable represented by ν:

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

Note here that to keep the model parsimonious, we do not allow the parameter τ to vary with ν.

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 adjusted R^2 statistic, resulting in:

$$\begin{aligned}\Delta Y_{it} = \alpha_{it} + \rho\Delta Y_{it-1} + \phi Y_{it-1} + \beta_U U_{it} + \beta_{U\nu}\,(\nu_{it}U_{it})& \\ + b_{01}\sum_Z F_{Zit} + b_{02}\sum_Z \nu_{it}F_{Zit} + b_{11}\sum_Z g_1(Z)F_{Zit}& \\ + b_{12}\sum_Z g_1(Z)\nu_{it}F_{Zit} + b_{21}\sum_Z g_2(Z)F_{Zit} + b_{22}\sum_Z g_2(Z)\nu_{it}F_{Zit} + \varepsilon_{it}&\end{aligned} \tag{3.6}$$

Armed with the parameters of function (3.5), we can evaluate the function at different levels of ν, 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, using an overlapping four-quarter window. These standard errors are consistent in the presence of heteroskedasticity, cross-correlation and autocorrelation of up to one year.[7]

Table IV estimates equation (3.6) separately for each quintile of market capitalization, replacing the intercept α with time dummies, and using the lagged level of institutional ownership (Y_{it-1}) in place of the interaction variable ν_{it}. 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 adjusted 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 adjusted R^2 statistics remain fairly modest, but this should not be surprising given the incentives that institutions have to conceal their activity using order-splitting strategies, and the increasing use of internalization and off-market matching of trades by institutional investors.

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

[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.

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. First, in an attempt to reduce transactions costs, institutional investors have increasingly adopted algorithmic trading strategies, such as volume-weighted average price (VWAP) engines. Such strategies result in large orders being broken up into smaller sizes, in an attempt by institutions to conceal their identity from the market-maker.[8] Second, institutions may use "scrum trades" to clean up their portfolios by closing out extremely small equity positions. Third, institutions may use extremely small "iceberg trades" to test the liquidity of the market before trading in larger sizes. Finally, it is possible 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 forecasting power of our method to the adjusted R^2 statistics implied by the LR method. The last row of Table II shows the implied adjusted R^2 statistics generated by out of sample forecasts from the Nelson-Siegel specification. These out of sample adjusted 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

[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.

each period, progressively forecasting one period ahead. Across all size quintiles of stocks, the resulting out of sample adjusted R^2 statistics are higher than either the restricted or unrestricted LR adjusted R^2 statistics.

4. Daily 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, the daily function is:

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

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 leads and lags within a quarter with all of the right hand side variables in equation (4.1). This exogeneity 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 therefore set the values of these parameters to zero when constructing daily flows. 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 market-adjusted stock returns and flows (demeaned by the daily cross-sectional mean return and mean flow in all cases), 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 adjusted R^2 statistic from Table II for each quintile. For example, for the median size quintile of stocks, we use the 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. Finally, we remove all stock-days for which flow observations cannot be computed due to non-availability of TAQ data.

There are several features of interest in Table V. First, for both types of our flows (but not for the LR flows), the means indicate that intra-quarter, institutions have been buying into large-cap stocks, and selling out of small and mid-cap stocks.[9] Interestingly, in our sample, daily 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. Second, median flows are generally greater than mean flows, with the exception of flows in the largest size 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.

Third, the standard deviations of our two flow measures are similar in magnitude, and, except for the smallest size quintile of stocks, always lower than the LR flow standard deviation. Fourth, 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-

[9] Note that these moments are of the fitted values from the daily function (4.1). Since we do not incorporate the quarterly parameters (ρ, ϕ, α and ε) when constructing daily flows, the moments will not necessarily match up with those in Table I.

bounce, as we can see from the large negative first daily autocorrelation of small stock returns. Fifth, 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 (except for the smallest size quintile of stocks) for our flow measures than for the LR measure.

Sixth, the contemporaneous daily correlation between the flow and return measures is high and positive for the three largest quintiles of stocks for all three flow measures. This suggests that investigating intra-day flow-return relationships may also yield interesting insights. 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, how do institutional flows at the daily frequency behave? Do past flows predict future flows? Do past returns predict future flows?

In order to answer these questions, we need a summary of the dynamics of flows and returns in terms of past movements in each and the other variable. We want to allow for higher-order dynamics, but unrestricted high-order VARs have too many free parameters to reliably estimate. Accordingly we model the dependence of flows and returns on one another as exponentially weighted moving averages (EWMA), writing for a stock i on a day d:

$$f_{id} = \alpha_f + \sum_{k=1}^{3} \phi_k^f f_{i,k,d-1} + \sum_{k=1}^{3} \phi_k^r r_{i,k,d-1} + \varepsilon_{id}^f \tag{4.2}$$

$$r_{id} = \alpha_r + \sum_{k=1}^{3} \rho_k^f f_{i,k,d-1} + \sum_{k=1}^{3} \rho_k^r r_{i,k,d-1} + \varepsilon_{id}^r \tag{4.3}$$

Here the left-hand side variables f_{id} and r_{id} are 'market-adjusted' by subtracting the daily cross-sectional mean flow and return across all stocks, respectively. The right hand side variables in each case are exponentially weighted moving averages of past flows and returns, with different half-lives. That is,

$$\begin{aligned} f_{i,k,d-1} &= (1-\chi_k) \sum_{j=1}^{d-1} \chi_k^j f_{i,d-j}, \\ r_{i,k,d-1} &= (1-\chi^k) \sum_{j=1}^{d-1} \chi_k^j r_{i,d-j}, \end{aligned}$$

and χ_k for $k = 1, 2, 3$ are set such that the half-life of a daily flow or return shock is $1, 10$, and 25 days respectively. We use these half-lives to incorporate both short- and long-run dynamics into our VAR specifications, and find that the likelihood function is flat with respect to minor variations in the χ_k parameters. We estimate the parameters ϕ and ρ using ordinary least squares.

We have estimated equations (4.2) and (4.3) using Lee-Radhakrishna (LR) flows and both our in-sample and out-of-sample flow measures. We find that the LR results differ in several important respects from the results generated by our two flow measures, which are quite similar to one another. These patterns are robust to skipping one day to avoid any potential contamination of our results by bid-ask bounce. 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.

Table VI summarizes the VAR dynamics for LR flows, and Table VII summarizes the dynamics for our in-sample flows. The top panel of each table reports the flow equation (4.2). The coefficients on lagged flows show that both the LR flows and our flows are persistent, but our flows are more so, consistent with the daily autocorrelations reported in Table V. In the market microstructure literature, persistent flows are generally thought to characterize the trading behavior of informed investors (Kyle (1985)). However, our measures represent changes in the *aggregate* ownership of institutional investors. Given this, persistence could also result from daily lead-lag effects across the trades of different institutional investors (Sias (2004), Hong, Kubik and Stein (2005)).

The coefficients of flows on lagged returns vary across size quintiles for LR flows in Table VI. They are all positive for the most rapidly decaying EWMA, but a mix of positive and negative for the more slowly decaying EWMAs with half-lives of 10 and 25 days. The results are much more consistent across size quintiles for our flows in Table VII. Here we see positive coefficients for the two more rapidly decaying EWMAs, with half-lives of 1 and 10 days, and negative coefficients for the slowly decaying EWMA with a half-life of 25 days. These results suggest that institutions are momentum traders in the short run, but contrarians in the longer run.

The bottom panels of Tables VI and VII report the coefficients for the return equation (4.3). The coefficients of returns on lagged returns, at the bottom of each panel, are predominantly negative for the EWMAs with 1-day and 10-day half-lives. This result 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 (Subrahmanyam (2005)).

The coefficients of returns on lagged flows are quite inconsistent across size quintiles for LR flows in Table VI, but a more consistent pattern appears for our flows in Table VII. We see that the most rapidly decaying EWMA of flows, with a 1-day half-life, has a significantly negative effect on future returns in all size quintiles except the fourth, while the most slowly decaying EWMA, with a 25-day half-life, has a significantly positive effect on future returns in all size quintiles except the third. This pattern suggests that institutions make short-term losses but longer-term profits on their trades.

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 (2007) who find that individual investor flows positively forecast returns at the weekly frequency. One explanation that is consistent with these findings can be found in Campbell, Grossman and Wang (1993), who model the interaction between groups of investors that have different 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 short-term stock return moving against institutional trading. The longer-term returns in the same direction as institutional trading suggest that even though institutions pay compensation for taking liquidity in the short-run, they have some longer-term ability to pick stocks.

In order to check whether daily returns respond differently to buying and selling pressure, in Table VIII we 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:

$$\begin{aligned} r_{id} &= \alpha_r + \sum_{k=1}^{3} \rho_k^b b_{i,k,d-1} + \sum_{k=1}^{3} \rho_k^s s_{i,k,d-1} + \sum_{k=1}^{3} \rho_k^r r_{i,k,d-1} + \varepsilon_{id}^r, \\ b_{i,k,d-1} &= f_{i,k,d-1}.I\{f_{i,k,d-1} > 0\}, \\ s_{k,d-1} &= -f_{i,k,d-1}.I\{f_{i,k,d-1} < 0\}. \end{aligned} \tag{4.4}$$

If the response of returns to institutional buying is identical to that of institutional selling, we should find that ρ_k^b and ρ_k^s have equal magnitudes and opposite signs.

Table VIII shows that in fact, the short-term liquidity effect of Table VII comes from institutional selling, not institutional buying. The response of returns to the EWMA of flows with a one-day half-life is positive whether the flows are buys or sells, indicating that institutions pay a much higher liquidity cost when they sell. The longer-term profits in Table VII also seem to come predominantly from the negative response of returns to the slowly decaying EWMA of sales with a 25-day half-life.

The asymmetry in short-term returns is consistent with the results of Kaniel, Saar, and Titman (2007) who find that buying by individual investors predicts positive returns, but 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 that institutional stock sales will consume more liquidity than institutional purchases.[10] 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. Institutional Flows and Earnings Announcements

5.1. Do institutions trade differently around earnings announcements?

We now use our methodology to understand the way in which institutions trade around earnings announcements. We measure earnings surprises in a way that is now standard in the literature, as deviations from mean predicted earnings scaled by stock price. Mean predicted earnings come from the Summary History file of the Institutional Brokers Estimate System (IBES).

[10] Kraus and Stoll (1972), Chan and Lakonishok (1993), Keim and Madhavan (1995), and Saar (2001) make similar arguments for asymmetry in the price impacts of block trades.

A preliminary issue is whether our basic regression for forecasting quarterly institutional ownership given daily trades of different sizes needs to be augmented to accommodate differences in institutional trading patterns around earnings announcements. We explore this issue by defining an event window around each earnings announcement and estimating regressions that allow the coefficients of the Nelson-Siegel regression to be different in the event window. We further allow the coefficients to depend linearly on the IBES earnings surprise. Appendix 4 summarizes the specifications that we consider. Only one of these appears to be a material improvement over the basic specification (3.6) that we used in the previous section. A specification that allows unclassifiable volume to have a different coefficient in the earnings announcement window, proportional to the earnings surprise, increases the adjusted R^2 statistic of the quarterly regression by about one percentage point for the smallest quintile of stocks, with a smaller increase for larger stocks. As a robustness check, we employ the daily institutional flows estimated using this augmented specification in several of our tests, but none of the qualitative patterns we report are sensitive to the exact specification we use.

5.2. Institutional trading and post-earnings-announcement drift

We begin by showing that post-earnings-announcement drift (PEAD) survives in our 1995-2000 sample period. We do this using the by-now standard methodology of Ball and Brown (1968), Foster, Olsen and Shevlin (1984), and Bernard and Thomas (1989): we sort stocks into equally weighted decile portfolios by the magnitude of normalized unexpected earnings on the announcement day, and examine the market-adjusted returns of each portfolio in a 120 day window around the earnings announcement.[11] PEAD refers to the tendency of the cumulative market-adjusted returns of stocks with high earnings surprises to drift up, while those of stocks with low earnings surprises drift down, in the 60 days following the announcement.

[11] In our benchmark specifications, we compute the earnings surprise as the reported earnings relative to the consensus (mean) forecast, normalized by the stock price. Our results are not materially affected by the measure of earnings surprise that we employ.

Figure 4 shows cumulative market-adjusted abnormal returns for each decile portfolio over the entire period from 60 days before to 60 days after the announcement. The left side of the figure shows the tendency for returns to anticipate IBES earnings surprises, while the right side shows the effect of PEAD. This effect is weaker in our sample period than in the earlier data of Bernard and Thomas (1989), which is not surprising considering the publicity that has been given to the PEAD phenomenon and the capital that has been deployed to exploit it. Nonetheless, we do see continued evidence of PEAD, particularly on the downside for stocks with disappointing earnings.

Next we inspect the behavior of our institutional flows for each of these portfolios. Figure 5 shows that the pattern of institutional flows in the event window is remarkably similar to the pattern followed by abnormal returns. In particular, institutions steadily buy stocks that experience positive earnings surprises, both before and after the announcement date. This implies that the complement of our institutional investors—small institutions and individual investors—must be selling these stocks.

Of course, we already know that institutions tend to buy stocks that have appreciated in the recent past, and the patterns shown in Figure 5 could follow mechanically from this tendency. In order to check that this is not driving the results, we replace our raw institutional flows with the residuals from the VAR system estimated in the previous section and plot the results in Figure 6. These residuals behave in much the same way as our raw flows, although their magnitude is smaller, and the patterns are somewhat less regular than in Figure 5.

In unreported results, we have also investigated whether this arbitrage-like behaviour of institutional flows varies with the characteristics of stocks. Institutions appear to be most aggressive in their earnings-window trading for mid-cap, high turnover stocks. We know that large stocks have the highest percentage of institutional ownership, and there may be fewer opportunities for institutions to make money from the apparent underreaction of unsophisticated investors in such stocks. The emphasis on high-turnover stocks is perhaps unsurprising, since liquid stocks are easier to arbitrage, and turnover is associated with liquidity (Amihud (2002)).

The portfolio approach provides a convenient graphical summary of PEAD, but we would like to know whether institutions trade to exploit this phenomenon at the level of individual stocks. In Table IX we forecast earnings surprises at the stock level, and in Table X, we forecast the post-earnings-announcement drift (the cumulative market-adjusted return in the 60 days following the earnings announcement) using cumulative flows in the 60 days before the earnings announcement. We control for size in the earnings surprise regression, and for size and pre-announcement stock returns in the PEAD regression. Each column in each of the tables uses a different measure of flows as the forecasting variable. The first column uses flows created using our basic specification (3.6), the second column uses a specification which allows the coefficient on unclassifiable volume to vary in proportion with the earnings surprise during the earnings announcement window (model 2 in Appendix 4), and the third column, entitled 'Cutoff rule' uses both small flows (created from trades less than $5,000 in size), and large flows calculated using the LR approach, with the best restricted cutoff rule for each size quintile chosen from the top panel of Table II.[12] The last three columns repeat the first three, but using residual flows from the VAR system estimated in the previous section. This control ensures that our surprise and drift forecasting regressions do not merely pick up return momentum around the time of the earnings announcement, a possibility given that flows are highly persistent, and tend to follow past return movements.

The results in Table IX show that the large LR flows appear to have some forecasting ability for the earnings surprise in our specifications, while small flows have no such predictive power. Institutional flows created using our method also have strong predictive ability for earnings surprises. However, when we substitute residual flows from the VAR specifications for all three flow measures, we find that only the flows created using our method continue to forecast the earnings surprise. The predictive power of large LR flows vanishes once we apply this stronger test.

[12] This is for consistency with Shanthikumar (2004) and Malmendier and Shanthikumar (2007), who use both small and large flows in their specifications.

Table X reveals very similar patterns from forecasting regressions for the post-earnings announcement drift. Here, however, the behaviour of the cutoff rule flows is different. Large LR flows now have no predictive power for the drift, while small flows positively forecast the drift. Institutional flows created using our method are also able to forecast the drift. Again, when we move to the specification that replaces flows with residual flows from the VAR, the forecasting ability of small flows does not survive. However the flows created using our method retain their forecasting power when faced with this tougher requirement.

Shanthikumar (2004) uses LR flows and small flows in the three-day window surrounding the earnings announcement to forecast the drift. She finds that LR flows have some negative forecasting power for the drift, and that small flows have positive forecasting power for the drift. While we find some forecasting power of small flows for the drift, residual small flows from a VAR specification have no such forecasting power. This suggests that any forecasting power of cutoff-rule based flows results from the persistence of these flows, and the tendency for flows to follow past return movements.

6. Conclusion

Much recent work has sought to infer high-frequency institutional trading behavior by cumulating large trades, on the grounds that these trades are more likely to be placed by institutions. In this paper we have shown that this straightforward approach does not give a good fit to the changes in institutional ownership measured each quarter from 13-F filings. We have developed an alternative approach that infers daily institutional trading from the coefficients of a regression of quarterly changes in institutional ownership on cumulative quarterly trades of different sizes.

Our method provides insights into the trading behavior of institutional investors. In our 1995–2000 sample period, 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, which is consistent with institutions employing algorithmic trading strategies such as VWAP engines. This is also consistent with institutions using these trades to 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.

Our regressions explain from 10 to 15% of the variance of quarterly changes in institutional ownership. This is a dramatic improvement over the naive approach of cumulating large trades, but the explanatory power of our approach remains modest. This should not be surprising given that institutions have incentives to avoid detection by intermediaries (Kyle (1985)) or by methods such as ours, and that they utilize order-splitting techniques to disguise their trades (Bertsimas and Lo (1998)). Furthermore, there has been an increasing use of internalization and off-market matching of trades by institutional investors, and there is considerable overlap between the trade sizes that may be used by wealthy individuals and by smaller institutions.

We use our method to provide new evidence on the daily dynamics of institutional flows and stock returns. We find that our measures of institutional flows predict short-term future daily stock returns negatively, a pattern that is not nearly as consistent when the naive approach is used to measure institutional flows. Most of this negative predictability is the result of institutional sales predicting high subsequent returns. Although consistent with the recent work of Kaniel, Saar and Titman (2007) on individual investors, our finding is surprising at first glance given the evidence for institutional outperformance at the quarterly frequency. 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). 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. Over longer periods, however, we find that institutions trade in a contrarian manner, buying stocks that have done poorly over the past few months. This pattern too shows up more clearly when we measure flows using our new method.

Finally, we use our method to show that institutional investors exploit the post-earnings-announcement drift. We first establish that traditional cutoff-rule based measures of institutional order flow have no ability to forecast earnings surprises or the post-earnings announcement drift once we control for the relationship between these flows, their own past values, and past return movements. In contrast, we find that institutional flows created using our method have robust predictive power for firm-level earnings surprises and the post-earnings-announcement drift.

7. Appendices

7.1. Appendix 1: Buy-Sell Classification

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

The analysis in Lee and Radhakrishna (2000) evaluates the effectiveness of the Lee and Ready matching algorithm, using the TORQ database, which has buy-sell classified, institutional-individual identified data for 144 stocks over a 3 month period. They find that after removing trades with potentially ambiguous classifications (such as trades that are batched or split up during execution), the buy/sell classification algorithm is 93 percent effective. In particular, they find that the accuracy is highest (at 98 percent) when trade-to-quote matching can be accomplished, lower (at 76 percent) for those trades that have to be classified using a tick test, and lowest (at 60 percent) for those trades classified using a zerotick test. We eliminate this last source of variability in our data by terming as unclassifiable those trades for which a zerotick test is required. We further identify as unclassifiable all trades that occur in the first half hour of trading (since these come from the opening auction) as well as any trade that is reported as cancelled, batched or split up in execution. This last category of trades is identified as unclassifiable since we use trade size as one important input into our prediction of institutional ownership. A trade that is reported as being batched or split up cannot be unambiguously classified in terms of its size. We aggregate all unclassifiable trades together, and use the bin of unclassifiable trades as an additional input into our prediction exercise.

7.2. 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.

7.3. 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 behavior 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.[13] 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.

7.4. Appendix 4: Institutional Trading Around Earnings Announcements

We consider a variety of models that allow the regression coefficients of our basic model to vary within an earnings announcement window and in proportion to the earnings surprise.

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

Our first model, which we call model 1, allows only the coefficient on unclassifiable volume to change in the earnings announcement window. Omitting the institutional ownership interaction for notational simplicity, the model is:

$$\begin{aligned}\Delta Y_{it} = \alpha + \rho \Delta Y_{i,t-1} + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_{U\nu} \left(\nu_{i,t-1} U_{it}\right) + \beta_{U^e} U^e_{it} \\ + b_{01} \sum_Z F_{Zit} + b_{02} \sum_Z \nu_{i,t-1} F_{Zit} \\ + b_{11} \sum_Z g_1(Z) F_{Zit} + b_{12} \sum_Z g_1(Z) \nu_{i,t-1} F_{Zit} \\ + b_{21} \sum_Z g_2(Z) F_{Zit} + b_{22} \sum_Z g_2(Z) \nu_{i,t-1} F_{Zit} + \varepsilon_{it} \quad (7.1)\end{aligned}$$

where the superscript e denotes data measured in the earnings announcement window.

A variant of this model, termed model 2, allows the coefficient on unclassifiable volume to vary in proportion with the earnings surprise during the earnings announcement window.

$$\begin{aligned}\Delta Y_{it} = \alpha + \rho \Delta Y_{i,t-1} + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_{U\nu} \left(\nu_{i,t-1} U_{it}\right) + \beta_{U^e \nu} U^e_{it} \nu_{it} \\ + b_{01} \sum_Z F_{Zit} + b_{02} \sum_Z \nu_{i,t-1} F_{Zit} \\ + b_{11} \sum_Z g_1(Z) F_{Zit} + b_{12} \sum_Z g_1(Z) \nu_{i,t-1} F_{Zit} \\ + b_{21} \sum_Z g_2(Z) F_{Zit} + b_{22} \sum_Z g_2(Z) \nu_{i,t-1} F_{Zit} + \varepsilon_{it} \quad (7.2)\end{aligned}$$

A second model, 3, which nests both of the previous models, allows both types of shift in the coefficient on unclassifiable volume.

$$\begin{aligned}\Delta Y_{it} = \alpha + \rho\Delta Y_{i,t-1} + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_{U\nu}\left(\nu_{i,t-1}U_{it}\right) + \beta_{U^e}U^e_{it} + \beta_{U^e\nu}U^e_{it}\nu_{it} \\ + b_{01}\sum_Z F_{Zit} + b_{02}\sum_Z \nu_{i,t-1}F_{Zit} \\ + b_{11}\sum_Z g_1(Z)F_{Zit} + b_{12}\sum_Z g_1(Z)\nu_{i,t-1}F_{Zit} \\ + b_{21}\sum_Z g_2(Z)F_{Zit} + b_{22}\sum_Z g_2(Z)\nu_{i,t-1}F_{Zit} + \varepsilon_{it} \quad (7.3)\end{aligned}$$

Model 4, which nests model 1, allows only unconditional differences in the coefficients on unclassifiable volume and the sum of all trade size bins. This amounts to an intercept shift in the Nelson-Siegel regression function.

$$\begin{aligned}\Delta Y_{it} = \alpha + \rho\Delta Y_{i,t-1} + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_{U\nu}\left(\nu_{i,t-1}U_{it}\right) + \beta_{U^e}U^e_{it} \\ + b_{01}\sum_Z F_{Zit} + b_{02}\sum_Z \nu_{i,t-1}F_{Zit} + b_{03}\sum_Z F^e_{Zit} \\ + b_{11}\sum_Z g_1(Z)F_{Zit} + b_{12}\sum_Z g_1(Z)\nu_{i,t-1}F_{Zit} \\ + b_{21}\sum_Z g_2(Z)F_{Zit} + b_{22}\sum_Z g_2(Z)\nu_{i,t-1}F_{Zit} + \varepsilon_{it} \quad (7.4)\end{aligned}$$

Model 5, which nests model 4, allows unconditional differences in unclassifiable volume and all trade size bins in the earnings announcement window.

$$\begin{aligned}\Delta Y_{it} = \alpha + \rho\Delta Y_{i,t-1} + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_{U\nu}\left(\nu_{i,t-1}U_{it}\right) + \beta_{U^e}U^e_{it} \\ + b_{01}\sum_Z F_{Zit} + b_{02}\sum_Z \nu_{i,t-1}F_{Zit} + b_{03}\sum_Z F^e_{Zit} \\ + b_{11}\sum_Z g_1(Z)F_{Zit} + b_{12}\sum_Z g_1(Z)\nu_{i,t-1}F_{Zit} + b_{13}\sum_Z g_1(Z)F^e_{Zit} \\ + b_{21}\sum_Z g_2(Z)F_{Zit} + b_{22}\sum_Z g_2(Z)\nu_{i,t-1}F_{Zit} + b_{23}\sum_Z g_2(Z)F^e_{Zit} + \varepsilon_{it} \quad (7.5)\end{aligned}$$

Model 6, which nests model 2, allows coefficients on unclassifiable volume and trade size bins to shift only in proportion to earnigns surprises. Note that the earnings surprise is not included on its own.

$$\begin{aligned}\Delta Y_{it} = \alpha + \rho\Delta Y_{i,t-1} + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_{U\nu}\left(\nu_{i,t-1}U_{it}\right) + \beta_{U^e\nu}U^e_{it}\nu_{it} \\ + b_{01}\sum_Z F_{Zit} + b_{02}\sum_Z \nu_{i,t-1}F_{Zit} + b_{04}\sum_Z F^e_{Zit}\nu_{it} \\ + b_{11}\sum_Z g_1(Z)F_{Zit} + b_{12}\sum_Z g_1(Z)\nu_{i,t-1}F_{Zit} + b_{14}\sum_Z g_1(Z)F^e_{Zit}\nu_{it} \\ + b_{21}\sum_Z g_2(Z)F_{Zit} + b_{22}\sum_Z g_2(Z)\nu_{i,t-1}F_{Zit} + b_{24}\sum_Z g_2(Z)F^e_{Zit}\nu_{it} + \varepsilon_{it}\end{aligned} \quad (7.6)$$

Model 7, which nests model 5, allows unconditional differences in all trade size bin coefficients, but conditional differences only in unclassifiable volume.

$$\begin{aligned}\Delta Y_{it} = \alpha + \rho\Delta Y_{i,t-1} + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_{U\nu}\left(\nu_{i,t-1}U_{it}\right) + \beta_{U^e}U^e_{it} + \beta_{U^e\nu}\nu_{it}U^e_{it} \\ + b_{01}\sum_Z F_{Zit} + b_{02}\sum_Z \nu_{i,t-1}F_{Zit} + b_{03}\sum_Z F^e_{Zit} \\ + b_{11}\sum_Z g_1(Z)F_{Zit} + b_{12}\sum_Z g_1(Z)\nu_{i,t-1}F_{Zit} + b_{13}\sum_Z g_1(Z)F^e_{Zit} \\ + b_{21}\sum_Z g_2(Z)F_{Zit} + b_{22}\sum_Z g_2(Z)\nu_{i,t-1}F_{Zit} + b_{23}\sum_Z g_2(Z)F^e_{Zit} + \varepsilon_{it}\end{aligned} \quad (7.7)$$

Finally, model 8 nests all the other models. This allows all conditional and unconditional differences, in unclassifiable volume as well as in bin-specific coefficients. Thus it includes separate coefficients for earnings announcement windows as well as interaction coefficients with earnings surprises.

$$\begin{aligned}
\Delta Y_{it} = \alpha + \rho \Delta Y_{i,t-1} + \phi Y_{i,t-1} + \beta_U U_{it} + \beta_{U\nu} \left(\nu_{i,t-1} U_{it}\right) + \beta_{U^e} U_{it}^e + \beta_{U^e \nu} \nu_{it} U_{it}^e \\
+ b_{01} \sum_Z F_{Zit} + b_{02} \sum_Z \nu_{i,t-1} F_{Zit} + b_{03} \sum_Z F_{Zit}^e + b_{04} \sum_Z F_{Zit}^e \nu_{it} \\
+ b_{11} \sum_Z g_1(Z) F_{Zit} + b_{12} \sum_Z g_1(Z) \nu_{i,t-1} F_{Zit} + b_{13} \sum_Z g_1(Z) F_{Zit}^e + b_{14} \sum_Z g_1(Z) F_{Zit}^e \nu_{it} \\
+ b_{21} \sum_Z g_2(Z) F_{Zit} + b_{22} \sum_Z g_2(Z) \nu_{i,t-1} F_{Zit} + b_{23} \sum_Z g_2(Z) F_{Zit}^e + b_{24} \sum_Z g_2(Z) F_{Zit}^e \nu_{it} + \varepsilon_{it}
\end{aligned} \tag{7.8}$$

Table A.1

Are There Changes in Institutional Trading Strategies Around Earnings Announcements?

This table presents comparisons of Adjusted R-squared values using model specifications described in Appendix 4. The top panel compares the Adjusted R-squared in each specification with all specifications that can be nested into it. The 'In-Sample' Adjusted R-squared is that of our benchmark Nelson-Siegel specification in Table IV. The bottom panel compares the Adjusted R-squared in all specifications with Adjusted R-squared of Model 2. P-values are computed using the Likelihood-Ratio test and are reported under the Adjusted R-squared values in the table.

	Small	**Q2**	**Q3**	**Q4**	**Large**	**All**
In-Sample Adjusted R-squared	0.1251	0.1232	0.1187	0.1251	0.1362	0.1142
Model 1: Adj R-squared	0.1250	0.1237	0.1187	0.1250	0.1362	0.1142
p-value in test with In-Sample	*0.6420*	*0.1232*	*0.4287*	*1.0000*	*0.4057*	*1.0000*
Model 2: Adj R-squared	0.1398	0.1295	0.1266	0.1280	0.1373	0.1227
p-value in test with In-Sample	***0.0000***	***0.0000***	***0.0000***	***0.0002***	***0.0187***	***0.0000***
Model 3: Adj R-squared	0.1397	0.1297	0.1267	0.1280	0.1374	0.1227
p-value in test with Model 1	***0.0000***	***0.0000***	***0.0000***	***0.0002***	***0.0186***	***0.0000***
p-value in test with Model 2	*1.0000*	*0.3509*	*0.5740*	*0.7309*	*0.5563*	*1.0000*
Model 4: Adj R-squared	0.1260	0.1250	0.1194	0.1256	0.1365	0.1152
p-value in test with Model 1	***0.0292***	***0.0123***	*0.0767*	*0.0692*	*0.1886*	***0.0000***
Model 5: Adj R-squared	0.1263	0.1257	0.1199	0.1267	0.1364	0.1158
p-value in test with Model 4	*0.3776*	*0.1587*	*0.1878*	*0.0588*	*1.0000*	*0.0008*
Model 6: Adj R-squared	0.1398	0.1293	0.1266	0.1291	0.1373	0.1227
p-value in test with Model 2	*0.6733*	*0.8843*	*0.6880*	*0.1043*	*0.7091*	*0.6524*
Model 7: Adj R-squared	0.1405	0.1315	0.1275	0.1294	0.1374	0.1239
p-value in test with Model 2	*0.3287*	*0.0278*	*0.2233*	*0.0987*	*0.6584*	*0.0000*
p-value in test with Model 5	***0.0000***	***0.0000***	***0.0000***	***0.0004***	***0.0186***	***0.0000***
Model 8: Adj R-squared	0.1407	0.1314	0.1275	0.1306	0.1374	0.1241
p-value in test with Model 2	*0.4479*	*0.1166*	*0.4114*	*0.0401*	*0.8392*	***0.0000***
p-value in test with Model 7	*0.5317*	*0.8839*	*0.6875*	*0.0758*	*0.7918*	*0.2406*

	Small	**Q2**	**Q3**	**Q4**	**Large**	**All**
Model 2: Adj R-squared	0.1398	0.1295	0.1266	0.1280	0.1373	0.1227
p-value in test with In-Sample	***0.0000***	***0.0000***	***0.0000***	***0.0002***	***0.0187***	***0.0000***
Model 3: Adj R-squared	0.1397	0.1297	0.1267	0.1280	0.1374	0.1227
p-value in test with Model 2	*1.0000*	*0.3509*	*0.5740*	*0.7309*	*0.5563*	*1.0000*
Model 6: Adj R-squared	0.1398	0.1293	0.1266	0.1291	0.1373	0.1227
p-value in test with Model 2	*0.6733*	*0.8843*	*0.6880*	*0.1043*	*0.7091*	*0.6524*
Model 7: Adj R-squared	0.1405	0.1315	0.1275	0.1294	0.1374	0.1239
p-value in test with Model 2	*0.3287*	***0.0278***	*0.2233*	*0.0987*	*0.6584*	***0.0000***
Model 8: Adj R-squared	0.1407	0.1314	0.1275	0.1306	0.1374	0.1241
p-value in test with Model 2	*0.4479*	*0.1166*	*0.4114*	***0.0401***	*0.8392*	***0.0000***

References

Acharya, V.V. and Lasse H. Pedersen, 2005, Asset pricing with liquidity risk, *Journal of Financial Economics*, 77, 375-410.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

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

Ball, Ray and P. Brown, 1968, An empirical evaluation of accounting income numbers, *Journal of Accounting Research* 6, 159-178.

Barber, Brad M. and Terrance Odean, 2000, Trading is hazardous to your wealth: the common stock investment performance of individual investors, *Journal of Finance* 55, 773–806

Barber, Brad M. and Terrance Odean, 2001, Boys will be boys: gender, overconfidence, and common stock investment, *Quarterly Journal of Economics* 116, 261–292.

Barber, Brad M. and Terrance Odean, 2007, All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors, forthcoming *Review of Financial Studies.*

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

Bartov, Eli, Suresh Radhakrishnan and Itzhak Krinsky, 2000, Investor sophistication and patterns in stock returns after earnings announcements, *The Accounting Review* 75, 43–64.

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.

Bernard, Victor, 1992, Stock price reactions to earnings announcements: a summary of recent anomalous evidence and possible explanations. *Advances in Behavioral Finance.* Russell Sage Foundation, New York.

Bernard, Victor L. and Jacob K. Thomas, 1989, Post-earnings-announcement drift: delayed price response or risk premium?, *Journal of Accounting Research* 27, 1–36.

Bertsimas, Dimitris and Andrew W. Lo, 1998. Optimal control of execution costs. *Journal of Financial Markets* 1,1–50.

Bhattacharya, Nilabhra, 2001, Investors' trade size and trading responses around earnings announcements: an empirical investigation, *The Accounting Review* 76, 221–244.

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

Brunnermeier, Markus K. and Stefan Nagel, 2004, Hedge funds and the technology bubble, *Journal of Finance*, 59(5), 2013–2040.

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.

Campbell, John Y., Tarun Ramadorai, and Tuomo O. Vuolteenaho, 2005, Caught on tape: Institutional order flow and stock returns, NBER Working Paper No. 11439.

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.

Chan, Louis K.C. and Josef Lakonishok, 1993, Institutional trades and intraday stock price behavior, *Journal of Financial Economics* 33, 173–199.

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.

Foster, George, Chris Olsen, and Terry Shevlin, 1984, Earnings releases, anomalies, and the behavior of security returns, *The Accounting Review* 59, 574–603.

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, 2007, Institutional portfolio flows and international investments, forthcoming *Review of Financial Studies*.

Froot, Kenneth A. and Melvyn Teo, 2007, Style investing and institutional investors, forthcoming *Journal of Financial and Quantitative Analysis.*

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.

Hirshleifer, David A., James N. Myers, Linda A. Myers, and Siew Hong Teoh, 2004, Do individual investors drive post-earnings announcement drift?, unpublished paper, Ohio State University.

Holthausen, Robert W., Richard W. Leftwich, and David Mayers, 1987, The effect of large block transactions on security prices: A cross-sectional analysis, *Journal of Financial Economics* 19, 237–267.

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, *Journal of Finance* 60, 2801–2824.

Hvidkjaer, Soeren, 2006, A trade-based analysis of momentum, *Review of Financial Studies* 19, 457–491.

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, 2007, Individual investor trading and stock returns, forthcoming *Journal of Finance*.

Ke, Bin and Santhosh Ramalingegowda, 2004, Do institutional investors exploit the post-earnings-announcement drift?, unpublished paper, Pennsylvania State University.

Keim, Donald B. and Ananth Madhavan, 1995, Anatomy of the trading process: empirical evidence on the behavior of institutional traders, *Journal of Financial Economics* 37, 371–398.

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.

Kraus, Alan and Hans W. Stoll, 1972, Price impacts of block trading on the New York Stock Exchange, *Journal of Finance* 27, 569–588.

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., 1992, Earnings news and small traders: an intraday analysis, *Journal of Accounting and Economics* 15, 265–302.

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.

Madhavan, Ananth and Minder Cheng, 1997, In search of liquidity: block trades in the upstairs and downstairs markets, *Review of Financial Studies* 10, 175–203.

Malmendier, Ulrike and Devin Shanthikumar, 2007, Are small investors naive about incentives?, *Journal of Financial Economics*, forthcoming.

Nagel, Stefan, 2005, Short sales, institutional investors, and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

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.

Odean, Terrance, 1998, Are investors reluctant to realize their losses?, *Journal of Finance* 53, 1775–1798.

Odean, Terrance, 1999, Do investors trade too much?, *American Economic Review* 89, 1279–1298.

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

Pastor, Lubos and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642-685.

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.

Saar, Gideon, 2001, Price impact asymmetry of block trades: an institutional trading explanation, *Review of Financial Studies* 14, 1153–1181.

Shanthikumar, Devin, 2004, Small and large trades around earnings announcements: does trading behavior explain post-earnings-announcement drift?, unpublished paper, Harvard Business School.

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, 2006, The price impact of institutional trading, *Journal of Business* 79, 2869–2910.

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 cutoff in each case are considered institutional. In the rows labeled 'restricted coefficients,' the coefficient on flows above the cutoff is constrained to be +1. In the 'free coefficients' specifications, the coefficient on flows above the cutoff is 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 from Table IV. The final row shows the one period ahead 'Out of Sample' adjusted R-squared of our regressions from an expanding window regression updated with one additional period each step, beginning with the first eight calendar quarters in the dataset.

Upper Cutoff *Adjusted R-Squared→*	**Small**	**Q2**	**Q3**	**Q4**	**Large**
Restricted Coefficient					
Cutoff = 5,000	-0.125	-0.124	-0.060	-0.003	0.038
Cutoff = 10,000	-0.079	-0.096	-0.036	0.008	0.045
Cutoff = 20,000	-0.044	-0.063	-0.008	0.023	0.053
Cutoff = 50,000	-0.012	-0.032	0.022	0.042	0.067
Cutoff = 100,000	-0.005	-0.020	0.038	0.052	0.079
Free Coefficient					
Cutoff = 5,000	0.069	0.041	0.058	0.076	0.097
Cutoff = 10,000	0.074	0.046	0.065	0.081	0.101
Cutoff = 20,000	0.076	0.053	0.074	0.086	0.105
Cutoff = 50,000	0.074	0.056	0.083	0.092	0.109
Cutoff = 100,000	0.067	0.053	0.084	0.091	0.111
In Sample Adjusted R-Squared	0.123	0.100	0.142	0.133	0.142
Out of Sample Adjusted R-Squared	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 (*Δ(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.

	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*
Δ(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
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 ('In-Sample Flows'), and flows constructed using out-of-sample estimated coefficients from our method ('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 daily by market capitalization (size). All days for which flow or return observations are missing are removed from the data before summary statistics are computed.

	Small	Q2	Q3	Q4	Large
Mean					
Lee-Radhakrishna Flows	-0.696	0.087	0.415	0.127	0.030
In-Sample Flows	-0.495	-0.088	-0.059	0.027	0.742
Out of Sample Flows	-0.601	-0.248	-0.281	0.272	0.650
Returns	-17.325	-6.381	-1.626	1.228	3.332
Median					
Lee-Radhakrishna Flows	-0.524	0.701	0.875	0.305	-0.167
In-Sample Flows	0.376	0.523	0.500	0.444	0.750
Out of Sample Flows	0.073	0.262	0.238	0.464	0.546
Returns	-22.201	-17.087	-12.958	-10.239	-6.903
Standard Deviation					
Lee-Radhakrishna Flows	2.193	10.655	12.688	12.674	9.362
In-Sample Flows	12.265	9.196	9.122	9.206	6.665
Out of Sample Flows	4.700	6.657	7.866	7.680	5.732
Returns	490.764	329.750	278.701	244.624	221.750
First Daily Autocorrelation					
Lee-Radhakrishna Flows	0.750	0.080	0.090	0.148	0.263
In-Sample Flows	0.122	0.162	0.173	0.187	0.247
Out of Sample Flows	0.196	0.165	0.175	0.204	0.253
Returns	-0.096	0.001	0.013	-0.010	-0.003
Corr(Flows(t),Returns(t))					
Lee-Radhakrishna Flows	0.028	0.093	0.138	0.203	0.309
In-Sample Flows	0.001	-0.055	0.072	0.227	0.315
Out of Sample Flows	-0.031	-0.063	0.046	0.189	0.298
Corr(In-Sample(t), LR(t))	0.219	0.672	0.835	0.900	0.913
Corr(Out of Sample(t), LR(t))	0.288	0.652	0.824	0.899	0.911
Corr(In-Sample(t), Out of Sample(t))	0.676	0.975	0.977	0.977	0.980

Table VI
Vector Autoregression of Daily Lee-Radhakrishna Flows and Returns

This table presents estimates of regressions of a VAR system (using exponentially weighted moving averages (EWMAs) of RHS variables) of daily Lee-Radhakrishna flows estimated using the best restricted cutoff rule specification for each size quintile chosen from Table II, and daily stock returns. Flows and returns are cross-sectionally demeaned each day to market adjust them. Flows are expressed in percentage points of market capitalization of the firm. We estimate the equations:

$$f_{i,d} = \alpha_f + \sum_{k=1}^{3} \phi_k^f f_{k,d-1} + \sum_{k=1}^{3} \phi_k^r r_{k,d-1} + \varepsilon_{i,d}^f$$

$$r_{i,d} = \alpha_r + \sum_{k=1}^{3} \rho_k^f f_{k,d-1} + \sum_{k=1}^{3} \rho_k^r r_{k,d-1} + \varepsilon_{i,d}^r$$

Here, $k = 1,2,3$ represent EWMAs of with half-lives of 1, 10 and 25 days respectively. Robust t-statistics computed using the Rogers (1983, 1993) method are reported in italics below the coefficients.

Lee-Radhakrishna Flow Equation	**Small**	**Q2**	**Q3**	**Q4**	**Large**
Flows(Half-Life 1 Day)	**0.045**	**0.066**	**0.094**	**0.143**	**0.194**
	7.060	*14.786*	*24.123*	*37.825*	*50.335*
Flows(Half-Life 10 Days)	**0.293**	**0.220**	**0.299**	**0.272**	**0.214**
	12.261	*12.359*	*18.230*	*16.188*	*11.794*
Flows(Half-Life 25 Days)	0.033	0.021	0.027	**0.151**	**0.325**
	1.436	*1.046*	*1.479*	*7.324*	*14.798*
Returns(Half-Life 1 Day)	**0.009**	**0.129**	**0.364**	**0.779**	**0.905**
	4.420	*10.219*	*18.421*	*31.127*	*42.581*
Returns(Half-Life 10 Days)	**0.027**	**0.369**	**0.476**	**-0.225**	**-0.477**
	3.306	*7.940*	*6.135*	*-1.966*	*-3.807*
Returns(Half-Life 25 Days)	**-0.012**	**-0.119**	-0.070	**0.412**	-0.106
	-2.736	*-3.767*	*-0.945*	*2.823*	*-0.581*
Adjusted R-squared	0.012	0.010	0.023	0.047	0.092
N(Observations)	798483	799051	798706	798493	798267
N(Firms)	1246	1490	1444	1262	774

Return Equation	**Small**	**Q2**	**Q3**	**Q4**	**Large**
Flows(Half-Life 1 Day)	**0.008**	-0.001	-0.001	**0.001**	-0.001
	1.797	*-1.507*	*-0.867*	*2.071*	*-1.480*
Flows(Half-Life 10 Days)	-0.015	**-0.008**	-0.001	**-0.009**	-0.007
	-0.967	*-2.318*	*-0.288*	*-3.442*	*-1.617*
Flows(Half-Life 25 Days)	**0.057**	0.000	**-0.009**	0.003	**0.010**
	3.978	*-0.054*	*-3.273*	*1.213*	*2.218*
Returns(Half-Life 1 Day)	**-0.332**	-0.010	**0.028**	**-0.024**	-0.004
	-31.044	*-0.955*	*3.898*	*-2.843*	*-0.328*
Returns(Half-Life 10 Days)	0.051	**-0.162**	**-0.169**	-0.002	**-0.148**
	1.187	*-6.286*	*-6.185*	*-0.065*	*-1.814*
Returns(Half-Life 25 Days)	**-0.044**	**0.077**	**0.105**	-0.029	0.108
	-1.923	*4.736*	*5.065*	*-0.704*	*0.884*
Adjusted R-squared	0.030	0.001	0.001	0.000	0.001
N(Observations)	798483	799051	798706	798493	798267
N(Firms)	1246	1490	1444	1262	774

Table VII
Vector Autoregression of Daily In-Sample Flows and Returns

This table presents estimates of regressions of a VAR system (using exponentially weighted moving averages (EWMAs) of RHS variables) of daily flows constructed using the coefficients we estimate in Table IV ('In-Sample Flows') and daily stock returns. Flows and returns are cross-sectionally demeaned each day to market adjust them. Flows are expressed in percentage points of market capitalization of the firm. We estimate the equations:

$$f_{i,d} = \alpha_f + \sum_{k=1}^{3} \phi_k^f f_{k,d-1} + \sum_{k=1}^{3} \phi_k^r r_{k,d-1} + \varepsilon_{i,d}^f$$

$$r_{i,d} = \alpha_r + \sum_{k=1}^{3} \rho_k^f f_{k,d-1} + \sum_{k=1}^{3} \rho_k^r r_{k,d-1} + \varepsilon_{i,d}^r$$

Here, $k = 1,2,3$ represent EWMAs of with half-lives of 1, 10 and 25 days respectively. Robust t-statistics computed using the Rogers (1983, 1993) method are reported in italics below the coefficients.

In-Sample Flow Equation	Small	Q2	Q3	Q4	Large
Flows(Half-Life 1 Day)	**0.145**	**0.154**	**0.150**	**0.175**	**0.206**
	27.900	*35.606*	*38.579*	*47.849*	*55.357*
Flows(Half-Life 10 Days)	**0.515**	**0.406**	**0.456**	**0.386**	**0.333**
	25.835	*23.527*	*29.829*	*24.629*	*21.256*
Flows(Half-Life 25 Days)	**0.094**	**0.136**	**0.058**	**0.154**	**0.232**
	4.578	*7.098*	*3.362*	*8.450*	*13.018*
Returns(Half-Life 1 Day)	0.003	**0.028**	**0.148**	**0.331**	**0.456**
	1.466	*3.875*	*12.973*	*24.294*	*36.639*
Returns(Half-Life 10 Days)	**0.079**	**0.637**	**0.970**	**0.433**	**0.214**
	8.729	*21.131*	*19.903*	*6.489*	*2.802*
Returns(Half-Life 25 Days)	**-0.031**	**-0.278**	**-0.424**	**-0.290**	**-0.610**
	-5.799	*-13.180*	*-8.858*	*-3.378*	*-5.343*
Adjusted R-squared	0.101	0.073	0.073	0.088	0.115
N(Observations)	798483	799051	798706	798493	798267
N(Firms)	1246	1490	1444	1262	774

Return Equation	Small	Q2	Q3	Q4	Large
Flows(Half-Life 1 Day)	**-0.021**	**-0.016**	**-0.006**	0.000	**-0.005**
	-3.471	*-9.003*	*-5.823*	*0.321*	*-3.542*
Flows(Half-Life 10 Days)	**-0.048**	0.003	0.005	**-0.014**	-0.010
	-2.063	*0.486*	*1.168*	*-3.191*	*-1.566*
Flows(Half-Life 25 Days)	**0.101**	**0.011**	-0.005	**0.010**	**0.012**
	4.874	*1.777*	*-1.179*	*2.185*	*1.852*
Returns(Half-Life 1 Day)	**-0.332**	-0.012	**0.028**	**-0.023**	-0.002
	-31.044	*-1.183*	*4.028*	*-2.697*	*-0.198*
Returns(Half-Life 10 Days)	0.057	**-0.156**	**-0.166**	0.001	**-0.146**
	1.324	*-6.025*	*-6.091*	*0.014*	*-1.786*
Returns(Half-Life 25 Days)	**-0.048**	**0.073**	**0.099**	-0.034	0.127
	-2.085	*4.470*	*4.781*	*-0.834*	*1.046*
Adjusted R-squared	0.030	0.001	0.001	0.000	0.001
N(Observations)	798483	799051	798706	798493	798267
N(Firms)	1246	1490	1444	1262	774

Table VIII
The Relationship Between Daily Returns and In-Sample Buys and Sells

This table presents estimates of regressions of daily returns on daily flows constructed using the coefficients we estimate in Table IV ('In-Sample Flows'). We sub-classify daily In-Sample flows into 'In-Sample buys' (net flows greater than zero, denoted by *b*) and 'In-Sample sells' (net flows less than zero, denoted by *s*), using exponentially weighted moving averages (EWMAs) of RHS variables. Flows and returns are cross-sectionally demeaned each day to market adjust them. Flows are expressed in percentage points of market capitalization of the firm. We estimate the equations:

$$r_{i,d} = \alpha_r + \sum_{k=1}^{3} \rho_k^b b_{k,d-1} + \sum_{k=1}^{3} \rho_k^s s_{k,d-1} + \sum_{k=1}^{3} \rho_k^r r_{k,d-1} + \varepsilon_{i,d}^r$$

Here, $k = 1,2,3$ represent EWMAs of with half-lives of 1, 10 and 25 days respectively. Robust t-statistics computed using the Rogers (1983, 1993) method are reported in italics below the coefficients.

Return Equation	**Small**	**Q2**	**Q3**	**Q4**	**Large**
Buys(Half-Life 1 Day)	**0.202**	**0.022**	**0.009**	**0.010**	0.002
	15.845	*6.629*	*4.890*	*6.356*	*0.751*
Sells(Half-Life 1 Day)	**0.101**	**0.034**	**0.014**	**0.007**	**0.014**
	13.032	*14.047*	*10.454*	*5.123*	*6.033*
Buys(Half-Life 10 Days)	**-0.218**	-0.015	**0.011**	-0.009	-0.002
	-4.704	*-1.257*	*1.695*	*-1.414*	*-0.303*
Sells(Half-Life 10 Days)	-0.004	-0.006	-0.001	**0.018**	**0.017**
	-0.156	*-0.693*	*-0.257*	*3.157*	*1.870*
Buys(Half-Life 25 Days)	0.043	**-0.022**	**-0.024**	0.006	0.012
	0.964	*-1.883*	*-3.474*	*0.982*	*1.526*
Sells(Half-Life 25 Days)	**-0.141**	**-0.033**	-0.007	**-0.012**	-0.004
	-5.938	*-4.266*	*-1.436*	*-2.019*	*-0.415*
Returns(Half-Life 1 Day)	**-0.338**	-0.013	**0.027**	**-0.025**	-0.004
	-31.404	*-1.294*	*3.811*	*-2.914*	*-0.312*
Returns(Half-Life 10 Days)	0.058	**-0.156**	**-0.166**	-0.001	**-0.148**
	1.328	*-6.011*	*-6.071*	*-0.014*	*-1.804*
Returns(Half-Life 25 Days)	**-0.048**	**0.073**	**0.099**	-0.035	0.128
	-2.059	*4.451*	*4.771*	*-0.875*	*1.052*
Adjusted R-squared	0.101	0.073	0.073	0.088	0.115
N(Observations)	798483	799051	798706	798493	798267
N(Firms)	1246	1490	1444	1262	774

Table IX
Institutional Flows and Earnings Surprises

This table presents estimates of forecasting regressions for the earnings surprise, using different institutional flow measures. The table reports estimates of:

$$s_{i,d} = \beta_1 \left(\sum_{j=1}^{60} f_{d-j} \right) + \beta_2 MCap_{i,d} + \upsilon_{i,d} .$$

f denotes flows which are used to forecast the earnings surprise *s* for a stock *i* on earnings announcement date *d*. The columns in the table show the specific daily flow measure employed: In-Sample, which are constructed using the coefficients in Table IV; Interacted, which are estimated using a Nelson-Siegel specification which allows the coefficient on unclassifiable volume to vary in proportion with the earnings surprise during the earnings announcement window (model 2 in Appendix 3); Cutoff rule, which comprise Lee-Radhakrishna flows estimated using the best restricted cutoff rule specification for each size quintile chosen from Table II and Small flows (aggregated, net buy-classified less sell-classified trades less than $5,000 in size). The fourth, fifth and sixth columns substitute residual flows for each flow measure. These are the residuals from a regression of each flow measure on lagged flows and lagged returns, the exact form of these regressions is shown in Tables VI and VII. Robust t-statistics computed using the Rogers (1983, 1993) method are reported in italics below the coefficients.

	Flows			**Residual Flows**		
Earnings Surprise	**In-Sample**	**Interacted**	**Cutoff Rule**	**In-Sample**	**Interacted**	**Cutoff Rule**
Intercept	**-0.001**	**-0.001**	**-0.001**	**-0.001**	**-0.001**	**-0.001**
	-19.630	*-19.686*	*-18.251*	*-18.916*	*-18.911*	*-18.908*
Cumulative Flows [-60,-1]	**0.020**	**0.020**		**0.038**	**0.036**	
	5.407	*5.136*		*2.827*	*2.629*	
Cumulative Lee-Radhakrishna Flows [-60,-1]			**0.009**			0.009
			2.959			*1.155*
Cumulative Small Flows [-60,-1]			0.021			0.064
			0.333			*0.258*
Market-Adjusted Returns [-30,-1]	**0.007**	**0.007**	**0.007**	**0.007**	**0.007**	**0.008**
	8.847	*8.865*	*9.304*	*9.098*	*9.105*	*9.345*
Market-Adjusted Returns [-60,-31]	**0.005**	**0.005**	**0.005**	**0.005**	**0.005**	**0.006**
	7.198	*7.264*	*7.513*	*7.969*	*7.992*	*8.258*
Market Capitalization ($ MM)	**0.011**	**0.011**	**0.012**	**0.013**	**0.013**	**0.013**
	7.680	*7.745*	*8.154*	*7.697*	*7.705*	*7.615*
Adjusted R-squared	0.028	0.028	0.027	0.027	0.027	0.026
N(Observations)	31,114	31,114	31,114	31,114	31,114	31,114
N(Firms)	2153	2153	2153	2153	2153	2153

Table X
Institutional Flows and the Post-Earnings Announcement Drift

This table presents estimates of forecasting regressions for the magnitude of the post-earnings announcement drift, using different institutional flow measures. The columns indicate the flow measure *f* employed to forecast the drift (cumulative market-adjusted returns) for a stock *i* on earnings announcement date *d*. The bottom panel reports estimates of:

$$\left(\sum_{j=1}^{60} m_{i,d+j}\right) = \beta_1\left(\sum_{j=1}^{60} f_{d-j}\right) + \beta_2\left(\sum_{j=1}^{30} m_{d-j}\right) + \beta_2\left(\sum_{j=31}^{60} m_{d-j}\right) + \beta_2 MCap_{i,d} + \upsilon_{i,d}$$

The columns indicate the flow measure *f* employed to forecast the cumulative market-adjusted return *m* for a stock *i* between the day after the earnings announcement *d+1* until 60 days after the announcement *d+60*. The columns in the table show the specific daily flow measure employed: In-Sample, which are constructed using the coefficients in Table IV; Interacted, which are estimated using a Nelson-Siegel specification which allows the coefficient on unclassifiable volume to vary in proportion with the earnings surprise during the earnings announcement window (model 2 in Appendix 3); Cutoff rule, which comprise Lee-Radhakrishna flows estimated using the best restricted cutoff rule specification for each size quintile chosen from Table II and Small flows (aggregated, net buy-classified less sell-classified trades less than $5,000 in size). The fourth, fifth and sixth columns substitute residual flows for each flow measure. These are the residuals from a regression of each flow measure on lagged flows and lagged returns, the exact form of these regressions is shown in Tables VI and VII. Robust t-statistics computed using the Rogers (1983, 1993) method are reported in italics below the coefficients.

	Flow Type			Residual Flow Type		
Drift [+1,+60]	**In-Sample**	**Interacted**	**Cutoff Rule**	**In-Sample**	**Interacted**	**Cutoff Rule**
Intercept	**-0.018**	**-0.018**	**-0.019**	**-0.019**	**-0.019**	**-0.019**
	-1.902	*-1.913*	*-2.001*	*-2.004*	*-2.003*	*-1.994*
Cumulative Flows [-60,-1]	**0.406**	**0.407**		**0.845**	**0.832**	
	2.664	*2.641*		*2.272*	*2.190*	
Cumulative Lee-Radhakrishna Flows [-60,-1]			0.032			-0.157
			0.436			*-0.682*
Cumulative Small Flows [-60,-1]			**3.946**			1.384
			2.235			*0.268*
Market-Adjusted Returns [-1,0]	**0.184**	**0.185**	**0.181**	**0.184**	**0.184**	**0.186**
	4.558	*4.557*	*4.332*	*4.493*	*4.494*	*4.434*
Market-Adjusted Returns [-30,-1]	-0.029	-0.028	-0.019	-0.024	-0.024	-0.018
	-1.581	*-1.570*	*-1.004*	*-1.295*	*-1.285*	*-0.943*
Market-Adjusted Returns [-60,-31]	-0.034	-0.034	-0.020	-0.026	-0.026	-0.019
	-0.802	*-0.794*	*-0.477*	*-0.614*	*-0.609*	*-0.453*
Market Capitalization ($ MM)	0.248	0.250	**0.277**	**0.281**	**0.281**	**0.281**
	1.623	*1.630*	*1.719*	*1.752*	*1.752*	*1.771*
Rsq	*0.004*	*0.004*	*0.004*	*0.003*	*0.003*	*0.003*
N(Observations)	31,114	31,114	31,114	31,114	31,114	31,114
N(Firms)	2285	2285	2285	2285	2285	2285

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).

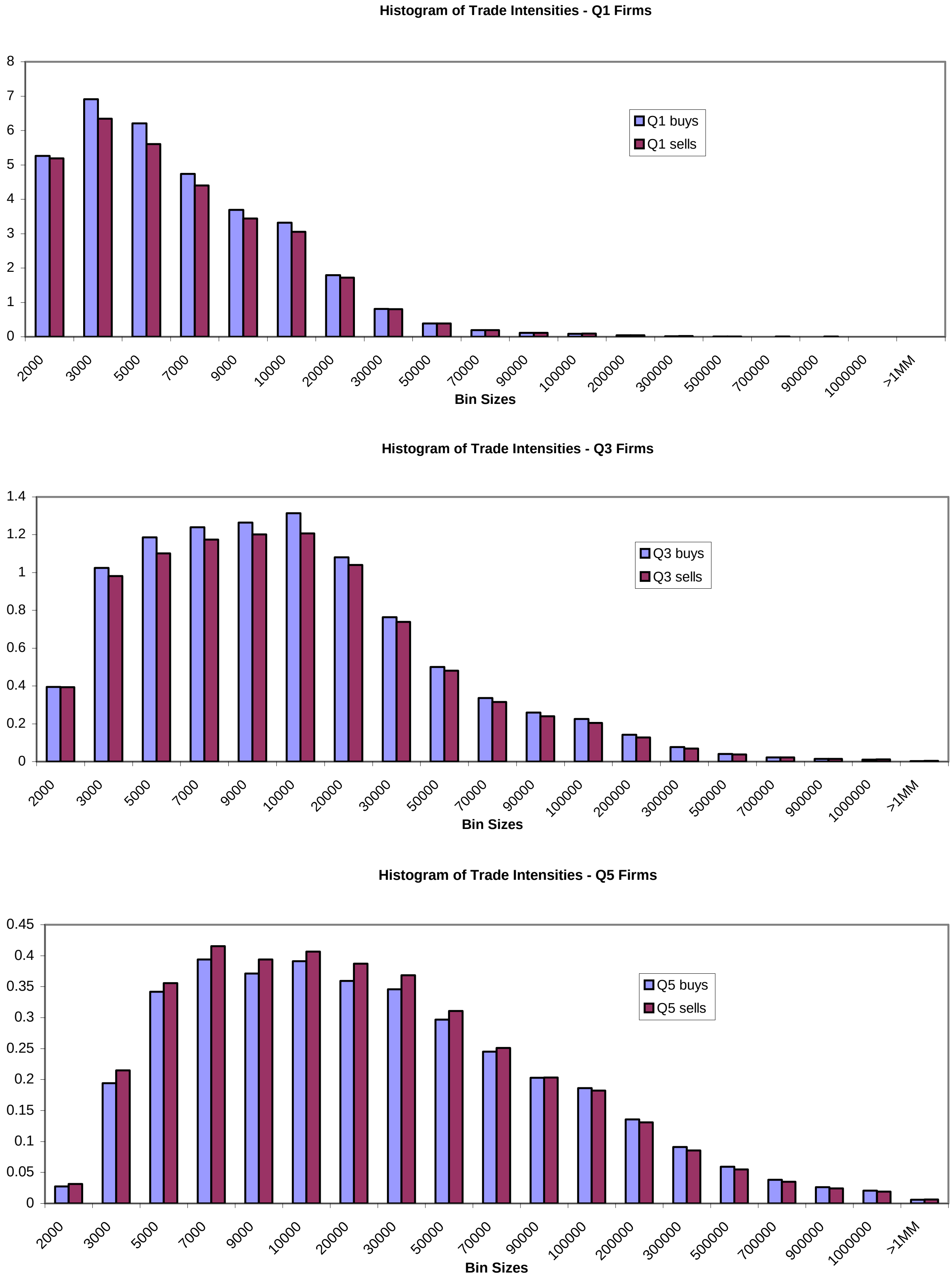

Figure 2

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.

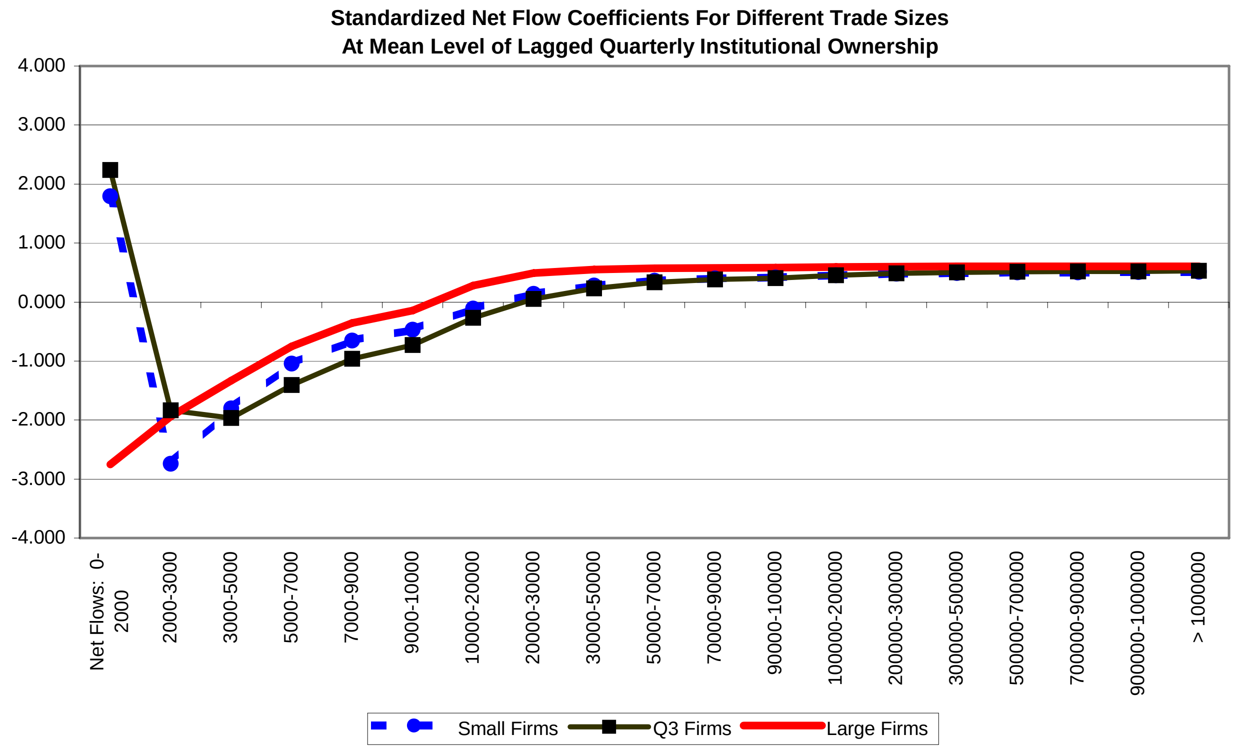

Figure 3

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 quarterly institutional ownership (LS) to its quarterly mean and to one standard deviation above its quarterly mean.

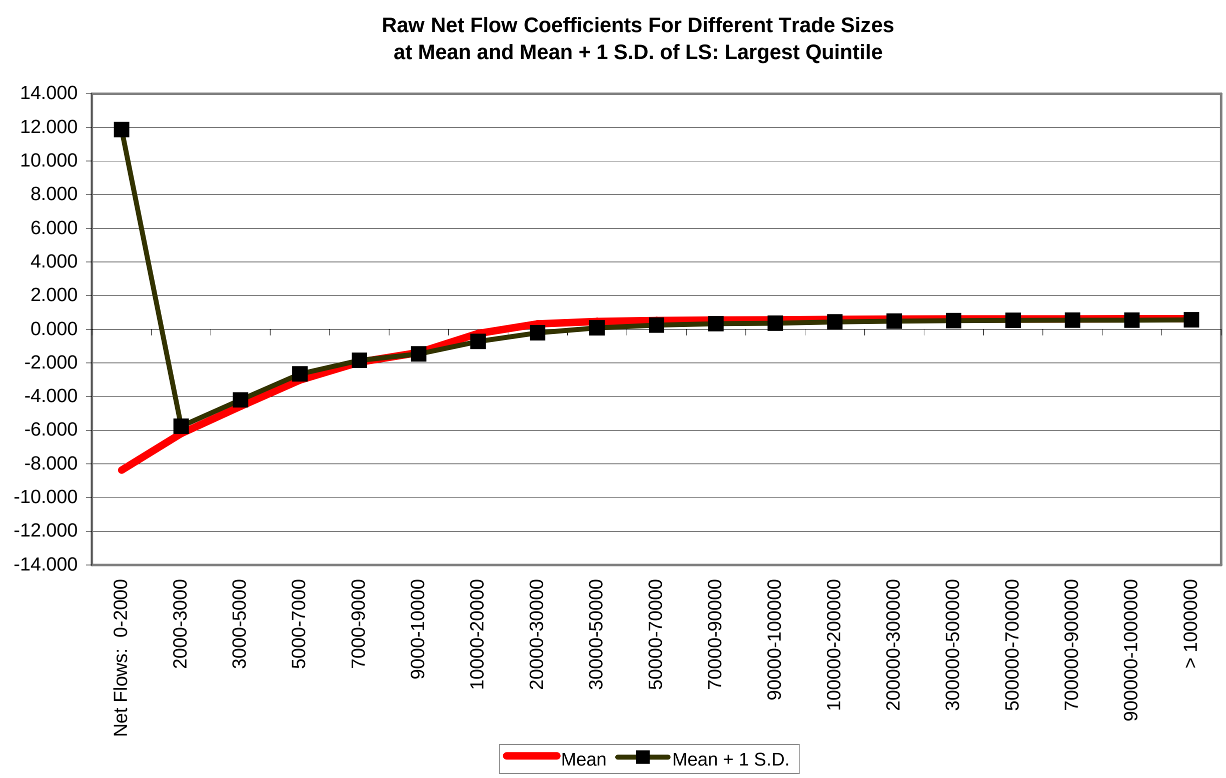

Figure 4

This figure plots cumulative abnormal stock returns computed using a market model (MARS), in each of ten standardized unexpected earnings (SUE) deciles. SUE's are computed relative to analyst mean forecasts. The figure shows cumulated stock returns in the entire window from [-60,+60]. P1 (P10) is the most negative (positive) earnings surprise portfolio.

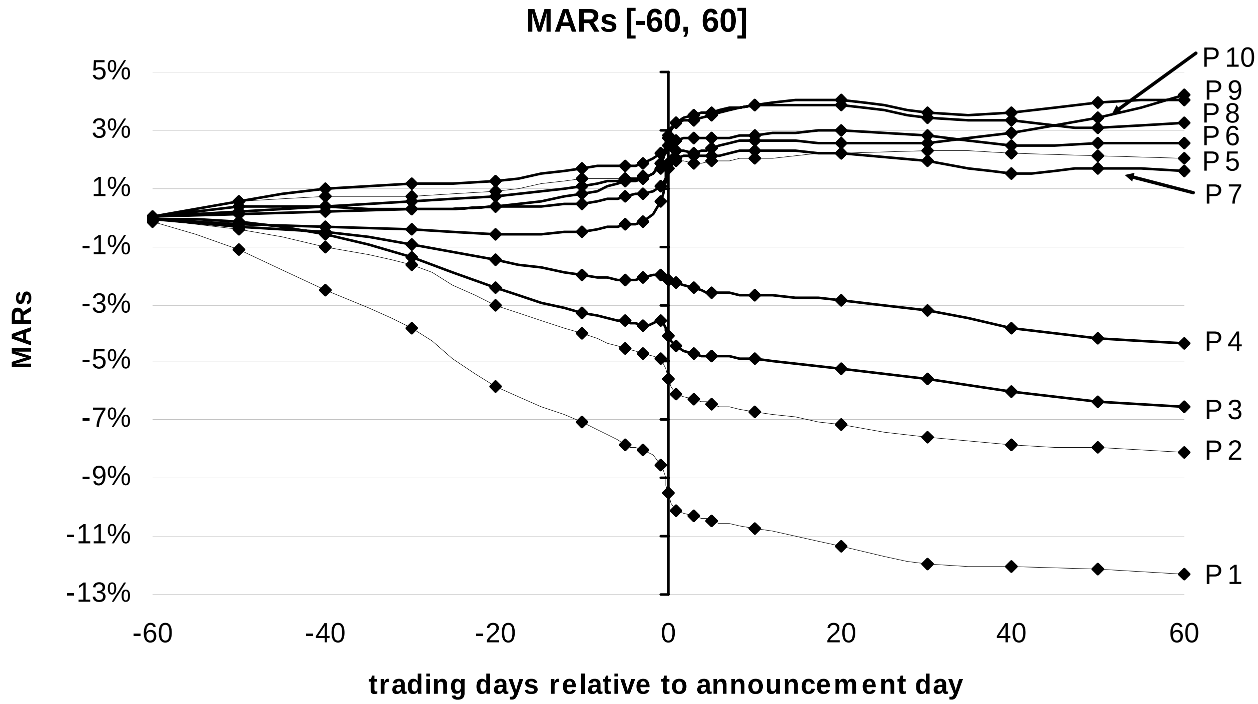

Figure 5

This figure plots cumulative institutional flows computed using a Nelson-Siegel specification which allows the coefficient on unclassifiable volume to vary in proportion with the earnings surprise during the earnings announcement window (model 2 in Appendix 3). The flows are aggregated into ten standardized unexpected earnings (SUE) deciles. SUE's are computed relative to analyst mean forecasts. The figure shows cumulated institutional flows as a percentage of daily market capitalization in the entire window from [-60,+60]. P1 (P10) is the most negative (positive) earnings surprise portfolio.

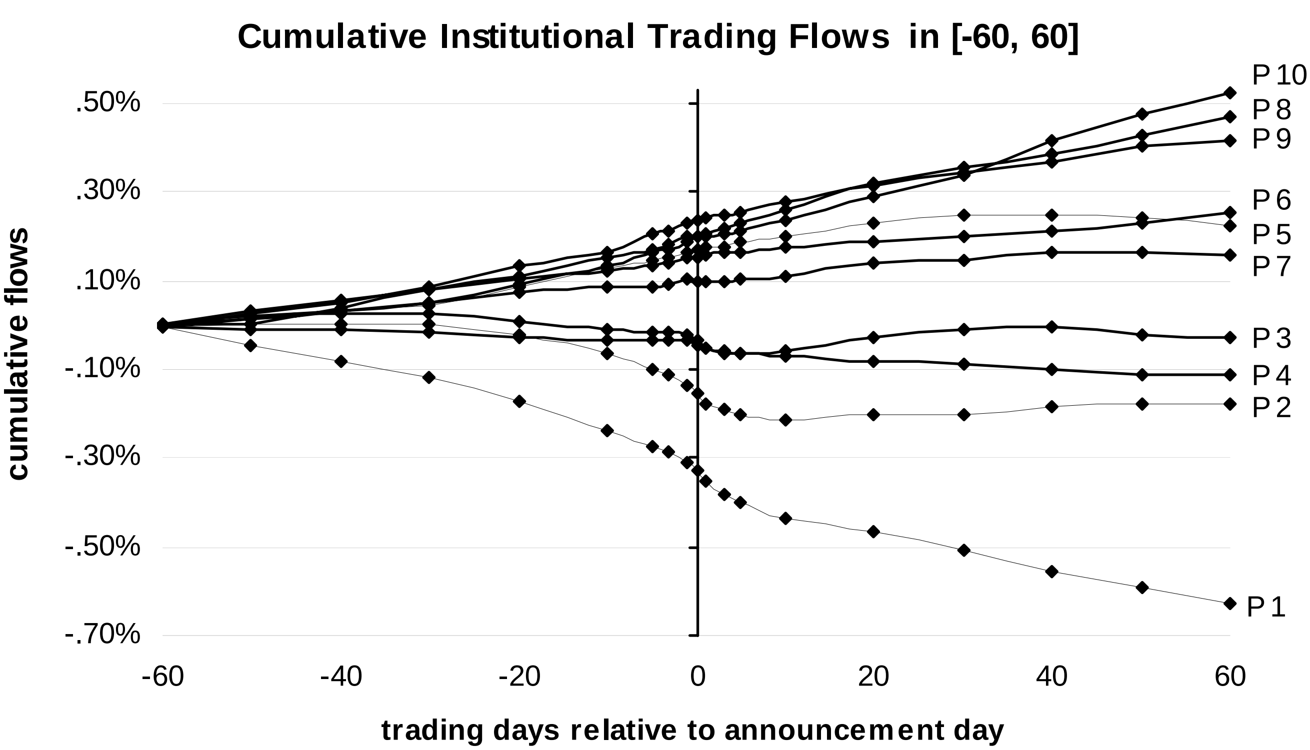

Figure 6

This figure plots cumulative residual institutional flows computed using a Nelson-Siegel specification which allows the coefficient on unclassifiable volume to vary in proportion with the earnings surprise during the earnings announcement window (model 2 in Appendix 3). The flows are the residuals from a regression of flows on past EWMA flows and returns as in Table VII. These residual flows are aggregated into ten standardized unexpected earnings (SUE) deciles. SUE's are computed relative to analyst mean forecasts. The figure shows cumulated institutional flows as a percentage of daily market capitalization in the entire window from [-60,+60]. P1 (P10) is the most negative (positive) earnings surprise portfolio