Market Microstructure and the Dynamic Relation of Stock Returns and Trading Flows

Terence Lim*

January 2002

Abstract

Market microstructure theories of trading predict that lead-lag relationship between stock returns and trading flows. I examine the dynamic relationship between daily returns and signed trading flow for a large sample of individual NYSE stocks between 1993 and 1999. The empirical evidence suggests that, at the forecast horizons studied in this paper, the reversal of stock returns following trading flow innovations is more consistent with the predictions of non-informational than with informational trading models. Also, the response of trading flows following return innovations suggests that feedback trading is more important at longer horizons. Economically, the magnitude of predictability is not large, but may be useful to investors interested in minimizing trading costs.

---

* This is a revised version of an earlier working paper titled "Market Microstructure and Stock Price Momentum" while the author was at Dartmouth College. I thank Robert Burnham and Boulat Minnigoulov of Dartmouth College for programming assistance with TAQ data, and Andrew Alford, Robert Jones and Arlen Khodadadi for helpful comments. Financial support from a Q-Group research grant is gratefully acknowledged.

1 Introduction

At the heart of market microstructure research are the questions of why investors trade and how price dynamics are related to trades with different motives. Investors may trade with other investors if they possess different beliefs or private information, or have unique investment opportunities or liquidity needs. If investors were homogeneous and the economy could be reduced to a single representative agent, then no trades take place at all, an implication noted by a series of "no trade" theorems such as Milgrom and Stokey (1982). Fortunately (for microstructure academicians and Wall Street brokers), investors are heterogeneous: their trading processes generate different dynamic relationships between trading flows and stock returns. Some investors may desire to buy or sell a stock for liquidity, or more generally, "non-informational" reasons: Trading then occurs between these and other investors to achieve risk-sharing. Investors may also trade for "informational" reasons: they may be better informed and attempt to profit from perceived misvaluations. Such speculative traders may pursue "value-motivated" or "positive feedback" trading strategies in response to stock price changes and their private information sets.

Noninformational and informational trades lead to different patterns of future stock price changes. For example, if a subset of investors desires to sell stock for noninformational reasons (such as an exogeneous change in risk aversion, private investment opportunities, or liquidity demands), other investors would be willing to accomodate the selling pressure only if the stock price drops such that they can be rewarded with high expected future returns. However, if a subset of investors sells a stock to speculate on private information, their selling pressure would be followed by low subsequent returns, as their negative information is eventually reflected in the stock price. In this example, noninformational selling pressure is followed by higher stock returns, while informational selling pressure is followed by low stock returns. More sophisticated theoretical models are developed in Campbell, Grossman and Wang (1993), Wang (1994), and Llorente et al (2000).

Investors who trade for speculative reasons may watch and respond to stock price deviations. We can loosely divide such speculative traders into two stylized groups, who respond with different patterns of trading flows following stock price innovations. "Value-motivated" traders, such as those modelled by Harris (1997), continuously compare prices to a private estimate of value. They sell (buy) aggressively when they believe prices have diverged above (below) their underlying fundamental values. Their trading flow appears contrarian – flows are negatively correlated with past returns – as they anticipate prices are likely to revert. It can be argued that such traders could earn rents and exist in capital markets because they stand ready to drive out inefficient price fluctuations. Gatev, Goetzmann and Rouwenhorst (1999) examined the profits available to a particular class of relative value arbitrageurs. On the other hand, positive feedback traders appear to chase prices – their trades are correlated with lagged stock returns. Some recent papers have theorized that such traders may prevalent in the capital markets. For example, Hong and Stein (1999) derive an economy with bounded rationality in which information diffuses only gradually, and hence such simple trend-chasing can be profititable. Alternatively, some investors may simply possess a behavioral tendency to extrapolate past trends, as suggested by Lakonishok, Shleifer and Vishny (1994).

Such motives for trading suggest different empirical implications for the dynamic relationship between trading flows and stock returns. I examine the daily flow-return dynamics of individual stocks traded on the NYSE, to assess the relative importance of these trading motives. Using tick-by-tick trades and quotes data (from TAQ), I infer the more aggressive side of each trade and compute daily estimates of the imbalance of buyer- and seller-initiated trading flows in each stock. Through daily cross-sectional regression analysis, I estimate lead-lag coefficients between net flows and returns across a large panel of individual stocks. I then calculate impulse response functions at various horizons and compare these to the predictions of the models. Stock prices initially rise (decline) but then appear to decline (rise) consistently for several weeks following positive (negative) trading flow innovations, which is more consistent with the predictions of non-informational than with informational trading models. The patterns of trading flow following return innovations suggest that feedback trading is more important at short horizons up to one week, but value-motivated trading at longer horizons.

Numerous empirical papers have investigated how the relation between return dynamics and trading *volume* depends on the extent of informational trading. Llorente et al (2000), Cooper (1999), Conrad, Hameed and Niden (1992), and Antoniewicz (1993) examined how return autocorrelations of individual stocks depend on trading volume. Stickel and Verrecchia (1994) examined if stock price responses after earnings announcements are more sustainable if accompanied by higher trading volume. However, models of trading make more direct predictions of relationships with (directional) trading *flows* than (non-directional) *volume*.

Some authors have also explored flow-return relationships using unique data sets that reveal the actual trading flows of a group of investors. Edelen and Warner (2000) study the relation between overall market returns and the aggregate flow into US equity mutual funds. Froot, O'Connell and Seasholes (2000) explore the behavior of international portfolio flows of institutional clients of a major custodial bank. However, such data sets are only available for a specific group of investors, and may be more appropriate for studying the behavior of these subsets of investors. This paper follows an approach of estimating daily flows from all available trades data, that is more representative of overall trading flows each day.

This work is related to Chordia and Subrahmanyam (2000), Chordia, Roll and Subrahmanyam (2000), Huang and Stoll (1994) and Hasbrouck (1991). Hasbrouck (1991) used a VAR framework to analyse the time-series dynamics of tick-to-tick trades and prices. Huang and Stoll draw on market microstructure theories for predicting high-frequency (5-minute) stock returns. I examine daily trade flows and stock returns. It is likely that most investors do not respond to tick-by-tick or high-frequency information, but tend to observe market data and react at lower (such as daily) frequencies. Furthermore, the heuristic algorithm used to infer trade direction is more accurate for aggregate than individual trades: see Lee and Radhakrishna (2000). Chordia and Subrahmanyam (2000) examined how daily and monthly expected stock returns can be explained by past order imbalances; other trading activity variables, such as volume and turnover; and firm characteristics. Chordia, Roll and Subrahmanyam (2000) pursued similar analysis on aggregate market indices. My analysis focuses on the joint lead-lag dynamics between returns and trading flows, and compare these relationships to the predictions of models of trading. I also examine longer term relationships (as supported by the data) up to 40 days, that may reverse shorter term relationships, since the trading horizons of different groups of investors may not be identical.

In section 2, I describe the data sample. Section 3 develops the empirical model used to examine the dynamic relation of individual stock returns and trading flows. Section 4 presents the key results as well as robustness tests. I conclude in section 5.

2 Data

I use tick-by-tick trades and quotes data from TAQ for the period 1993-1999. I included stocks with a primary listing on the NYSE because I use an algorithm for estimating the direction and magnitude of trade flows that was developed and tested by prior researchers principally for NYSE transactions. This was merged by CUSIP identifier with other stocks data from the Daily CRSP files, such as splits and dividends, share type, and shares outstanding. I excluded securities from my final sample that were not common stocks (CRSP share type of 10 or 11) or priced below five dollars.

Trades data were then purged for the following reasons: out of sequence; recorded before the open or after the closing time; or had special settlement conditions. Quotes data were also purged if: non-BBO (best bid- or offer-eligible); established before the opening of the market or after the close; or had negative bid-ask spreads. I then assigned trades as being either buyer- or seller-initiated following the midquote- and tick-test algorithm of Lee and Ready (1991): if a transaction occurs above the prevailing mid-quote (any quote less than five seconds prior to the trade is ignored and the first one at least five seconds prior is used), it is a purchase and vice versa. If a transaction occurs exactly at the mid-quote, it is signed using the previous transaction price according to the tick test (i.e. buyer-initiated if the sign of the last price change is positive and vice versa). Lee and Radhakrishna (2000) evaluated the accuracy of this algorithm for inferring order flow from publicly-available tick data. Validating with a small unique data set (TORQ) that identifies the true initiators of orders, they confirmed that the Lee and Ready algorithm is useful as a general technique for identifying the more aggressive side of a trade. However, for a significant fraction of trades (some 40 percent of their sample), neither side can even be sensibly labelled as its initiator. Thus they "caution researchers against using it to identify small samples of individual trades as clear buys or sells". The algorithm is more appropriate for inferring the net imbalances of large samples of trades.

Each day, and for each eligible stock, I computed the following variables:

FLOW : the estimated daily sum of buyer-initiated shares minus seller-initiated shares traded, scaled by total shares outstanding.

TURN : daily shares traded, scaled by shares outstanding.

RET : daily stock return (continuously-compounded), computed from closing mid-quotes ,adjusted for any stock splits or dividends.

Table 1 summarizes some descriptive statistics of the sample. On average, there were almost 1600 stocks throughout the period, slightly more in the latter half. On average, daily turnover was about 0.31 percent of shares outstanding, but this was higher in the latter half of the period and for larger stocks. Our (signed) FLOW measure was slightly positive (about plus 0.5 basis points of shares oustanding), due primarily to larger firms. The average cross-sectional standard deviation of FLOW was about 9.4 basis points of shares outstanding; this was slightly higher for small firms, indicating that despite lower average turnover, imbalances of trading flows were greater for smaller firms. FLOW and RET measures exhibited strong positive contemperaneous correlation: the average cross-sectional standard deviation was almost 33 percent,

[INSERT TABLE 1 HERE]

3 The Empirical Model

Let RET_{it} and $FLOW_{it}$,the return and trading flow respectively for security i on day t, be related by the following bivariate set of autoregressive equations:

$$RET_{i,t} - \overline{RET_i} = \sum_{j=1}^{j=J} a_j (RET_{i,t-j} - \overline{RET_i}) + \sum_{j=1}^{j=J} b_j (FLOW_{i,t-j} - \overline{FLOW_i}) + u_{i,t}$$

$$FLOW_{i,t} - \overline{FLOW_i} = \sum_{j=1}^{j=J} c_j (RET_{i,t-j} - \overline{RET_i}) + \sum_{j=1}^{j=J} d_j (FLOW_{i,t-j} - \overline{FLOW_i}) + v_{i,t}$$

A natural way to examine this relationship between daily stock returns and trading flows with their lagged values might be to separately estimate time-series regressions or, more generally, vector autoregressions for each security. However, the use of parameter estimates averaged across securities would pose a problem for statistical inference because of the cross-sectional dependence of the estimates. Instead, I fit the regression model cross-sectionally across stocks each day. The parameter estimates and standard errors are then obtained from the time-series of daily regression estimates, as in Jegadeesh (1990) and Fama and MacBeth (1973).

I found that the data supported using a lag length (J) of up to 40 days. $\overline{RET_i}$ and $\overline{FLOW_i}$ are unbiased estimates of the unconditional expected daily returns and trading flows of security i, which I estimated for each stock using the full sample period available: results were not sensitive to the choice of a shorter sample period.

We are interested in describing the response of a variable over time (at $t+1$, $t+2$, ...), due to an impulse in another variable at day t, but with all other variables dated t or earlier held constant. I compute impulse response functions from the autoregressive coefficient estimates to help answer this question, following standard treatments of Vector Autoregression equations such as chapter 11 of Hamilton (1994).

Let ϕ_j be the vector of jth-lag coefficients a_j, b_j, c_j, and d_j; and y_t be the vector of time t variables RET_t and $FLOW_t$, with mean μ. Then the autoregressive equations can be expressed as

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + ... + \phi_p(y_{t-p} - \mu) + \epsilon_t$$

We would like to re-write this in moving-average representation, where ϵ_t is the fundamental innovation for y_t.[1] The coefficient ψ_j from this infinite-MA representation measures the response j-periods ahead due to an impulse to the system at time t. [2]

$$y_t = \mu + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + ...$$

[1]From an econometric perspective, this requires an assumption that the variables are jointly covariance stationary, and that the distribution of the disturbances be uncorrelated with the regressors.

[2]The solution for ψ is given by:

$$\psi_1 - \phi_1 = 0$$

$$\psi_2 = \phi_1\psi_1 + \phi_2$$

$$\psi_s = \phi_1\psi_{s-1} + \phi_2\psi_{s-2} + ... + \phi_p\psi_{s-p}$$

with $\psi_0 = I$, $\psi_s = 0$ for $s < 0$.

Standard errors for the impulse response coefficients, ψ_s, are derived by Monte Carlo simulation. Let $\hat{\pi}$ and $\hat{\Omega}$ be the time-series mean and covariance matrix of the daily cross-sectional regression coefficient estimates a, b, c, d. Randomly generate a vector $\pi^{(1)}$ drawn from a $N(\hat{\pi}, \frac{1}{t}\hat{\Omega})$ distribution, and calculate $\psi_s(\pi^{(1)})$. Draw a second vector $\pi^{(2)}$ from the same distribution and calculate $\psi_s(\pi^{(2)})$. This process is repeated 1000 times, producing a simulated distribution for the estimate of ψ_s.

4 Empirical Results

4.1 Bivariate Regressions

I fit the bivariate regression equations cross-sectionally each day. Observations where stock price was less than five dollars were excluded. To reduce the impact of outliers, I dropped observations each day where either FLOW or RET were in the tail one percent of values, and truncated regressors at their tail 2.5 percent values. In table 2, I report the mean coefficient estimates as well as cumulative impulse response coefficients at various horizons. Figure 1 plots the cumulative response functions out to 60 days, with 95 percent confidence bands, of stock returns following a one percent shock to trading flows in a day, and vice versa.

[INSERT TABLE 2 HERE]

[INSERT FIGURE 1 HERE]

Stock returns appear to be slightly positive the day following positive trading flows: the first lagged coefficient on FLOW for forecasting RET is +0.0591. However, subsequent lagged coefficients turn negative, indicating that prices subsequently decline. The cumulative impulse response coefficients for 5-, 10-, 20- and 40-days indicate that returns continue to be negative for two months. This suggests that trading flows have a short initial price impact, but returns are subsequently lower (after positive flows) which is more consistent with the predictions of noninformational trading. Chordia and Subrahmanyam (2000) also find that order imbalances are positively related to returns the next day, but negatively to next month's returns.

After a positive one-day stock return innovation, trading flows appear to be positive for several days (particularly the first day). A week later, however, trading flows turn negative. This could be interpreted that positive-feedback trading is prevalent initially, but value-motivated trading becomes increasingly important after several days. The initial cumulative response of +0.0030 at five days, subsequently turns negative to –0.0016 and –0.0074 after 20- and 40-days respectively.

I also find that the own-autocorrelations of returns are negative beyond two days, consistent with the weekly return reversals found by prior researchers such as Lo and MacKinlay (1990), even when stock prices are computed from midquotes (which are free from bid-ask bounce biases). However daily midquote returns exhibit weak continuation, as was also reported by Chordia and Subrahmanyam (2000),

Trading flows exhibited strong positive auto-correlations at all horizons. This persistence in trading flows could be symptomatic of herding behavior, which may have little or no information content. Prior researchers, such as Nofsinger and Sias (1998), Wermers (1998) and Grinblatt, Titman and Wermers (1995), have found evidence of herding in the portfolio holdings of significant classes of investors,

The results presented exhibit strong significance *statistically*. These inferences are drawn from a very large sample. With about 1600 securities per day over almost 7 years, we have over 2 million observations. Consequently, conventional levels of significance may not be appropriate. We would also like to assess the *economic* importance of the relationships.

The cumulative response of return at 20 days following an innovation to trading flows is -0.3789. In other words, an innovation to trading flow of 10 basis points of shares outstanding (corresponding approximately to the value of one standard deviation of FLOW) results in a cumulative return of -

3.7 basis points over one month. The cumulative impulse response of FLOW following an innovation in RET is 0.0030 after five days, but –0.0016 after 20-days. Thus an innovation to returns of 1.6 percent (the value of one standard deviation cross-sectionally of RET) leads to cumulative trading flows of plus 0.5 basis points (of shares outstanding) after five days, but minus 0.25 basis points after 20 days. Recall that the daily standard deviation of flow is about 10 basis points. Hence overall, the economic significance of the relationship is not very large.

4.2 Robustness in subperiods

To evaluate the sensitivity of the results, I divided the sample into two time periods. Prior research such as Hvidkjaer (2000) suggest that trading around the turn-of-year may be strongly influenced by tax-related selling, hence I also considered a sample that excluded observations in the months of December, January and February. Table 3 summarizes results when evaluated over these subperiods.

[INSERT TABLE 3 HERE]

I find that estimated relationships are similar in both sample periods. Dropping the three months of observations that may be influenced by tax-related trading also made little difference.

4.3 Robustness in size-sorted subsamples

I next divided the sample by market capitalization each day into large and small firms. Table 4 summarizes results when evaluated over these size-sorted subsamples.

[INSERT TABLE 4 HERE]

The negative cumulative response of trading flows following return innovations for small stocks appears stronger than for large stocks. The initial response of flows to returns for large stocks is more positive (the one-day coefficient is +0.0013 for small stocks versus +0.0037 for large stocks), and the cumulative response function continues to rise out to five days (coefficient of +0.0061) before starting to decline. Value-motivated trading appears to be relatively more intense in small stocks. Other patterns remain qualitatively the same.

4.4 Univariate Lagged Regressions

In table 5, I report results from fitting univariate regressions, where lagged values for only one variable at a time are included as regressors. As before, daily cross-sectional regressions are fitted and inferences drawn from the time-series mean and standard errors of the estimates. The response of stock returns following innovations in either returns or trading flows is similar to the bivariate regression model. However, the relation of cumulative flows to past returns remains positive at all lags, whereas this relationship turns negative at longer lags in the bivariate model. This could be due to the strong positive relationship between FLOW and lagged FLOW, and between contemporaneous FLOW and RET. When these relationships are controlled for in a bivariate regression, the negative correlations between flows and a pure shocks to stock returns at longer lags are revealed.

[INSERT TABLE 5 HERE]

4.5 Time Series Bivariate Regressions

As an additional test of robustness, I estimated the bivariate relationships from time-series regressions on individual stocks. The sample included those 1157 stocks with at least 5 years of data available. Because of cross-sectional dependence of the estimates, I do not calculate statistical tests, but simply report the means and proportions of positive estimates. The observed lead-lag relationships between FLOW and RET exhibit qualitatively similar patterns to those previously obtained using the cross-sectional regression approach.

[INSERT TABLE 6 HERE]

5 Conclusion

Market microstructure theories that model the heterogeneity of investors suggest that predictable lead-lag relationships exist between stock returns and trading flows. The relative importance of these models of trading and sources of investor heterogeneity may be observed in the dynamic returns-flows relation.

Following a positive one-day shock to trading flows, stock prices initially rise the first day, but then decline after the second day. The lower subsequent returns are more consistent with the predictions of non-informational trading. Trading flows are positive for several days following a positive one-day return innovation, but then subsequently turn negative. This suggests that positive feedback trading is initially important but value-motivated trading becomes relatively more prevalent after a few days. These empirical findings appear robust in various subsamples and periods, and using alternative estimation approaches.

That predictable dynamic relationships exist between prices and trading flows, as suggested by market microstructure models, need not be inconsistent with an efficient market. The economic magnitudes are not large, and potential arbitrage profits after transactions costs may not be available to most investors. However, this evidence of predictability can be useful to investors interested in minimizing trading costs.[3]

[3] For example, forecasts of short-term trading flows and stock returns may be incorporated into estimates of price impact and slippage, to drive strategies for working large orders in order to reduce expected implementation costs.

References

Antoneiwicz, R. L., 1993, “Relative Volume and Subsequent Stock Price Movements, working paper, Board of Governors of the Federal Reserve System.

Campbell, John Y., Sanford J. Grossman, and Jiang Wang, 1993, “Trading Volume and Serial Correlation in Stock Returns”, *Quarterly Journal of Economics*, 905-939.

Chordia, Tarun, and Avanidhar Subrahmanyam, 2000, “Order Imbalance and the Cross-section of Expected Stock Returns”, working paper, UCLA.

Chordia, Tarun, Richard Roll and Avanidhar Subrahmanyam, 2000, “Order Imbalance, Liquidity and Market Returns”, working paper, UCLA.

Cooper, Michael, 1999, “Filter Rules Based on Price and Volume in Individual Security Overreaction”, *Review of Financial Studies* 12, 901-935.

Conrad, J. A. Hameed and Niden C. M., 1991, “Components of Short-Horizon Individual Security Returns”, *Journal of Finance* 49, 1305-1329.

Edelen, Roger and Jerold Warner, 2000, “Aggregate Price Effects of Institutional Trading: A Study of Mutual Fund Flows and Market Returns”, working paper, University of Pennsylvania.

Fama, Eugene, and J. D. MacBeth, 1973, “Risk, Return and Equilibrium:Empirical Test”, *Journal of Political Economy* 98, 247-273.

Gatev, Evan G., William N. Goetzmann and K. Geert Rouwenhorst, 1999, ”Pairs Trading: Performance of a Relative Value Arbitrage Rule”, NBER working paper 7032

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.

Hamilton, James. D, 1994, “Time Series Analysis”, Princeton University Press, Princeton, New Jersey.

Harris, Lawrence, 1997, “Optimal Dynamic Order Submission Strategies in Some Stylized Trading Problems”, working paper, USC.

Hasbrouck, Joel, 1991, “Measuring the Information Content of Stock Trades”, *Journal of Finance* 46, 179-207.

Hong, Harrison, and Jeremy C. Stein, 1999, “A Unified Theory of Over- and Under-reaction”, *Journal of Finance.*

Huang, Roger D., and Hans R. Stoll, 1994, “Market Microstructure and Stock Return Predictions”, *Review of Financial Studies* 7, 179-213.

Hvidkjaer, Soren, 2000, “A Trade-based Analysis of Momentum”, Cornell University working paper.

Jegadeesh, Narasimhan, 1990, “Evidence of Predictable Behavior of Security Returns”, *Journal of Finance* 55, 881-898.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, "Contrarian Investment, Extrapolation, and Risk, *Journal of Finance* 49, 1541-1578.

Lee, Charles, and Mark Ready, 1991, "Inferring trade direction from intraday data", *Journal of Finance* 46, 733-747.

Lee, Charles, and B. Radhakrishna, 2000, "Inferring Investor Behavior: Evidence from TORQ Data", *Journal of Financial Markets* 3, 83-111.

Llorente, Guillermo, Roni Michaely, Gideon Saar and Jiang Wang, 2000, "Dynamic Volume-Return Relation of Individual Stocks", Cornell University working paper.

Lo, Andrew W., and A. Craig MacKinlay, 1990, "When Are Contrarian Profits Due to Stock Market Overreaction?", *Review of Financial Studies* 3, 175-208.

Milgrom, P. and N. Stokey, 1982, "Information, Trade and Common Knowledge", *Journal of Economic Theory* 26, 17-27.

Nofsinger, John R. and Richard W. Sias, 1998, "Herding and Feedback Trading by Institutional and Individual Investors, working paper.

Stickel, Scott E., and Robert E. Verrecchia, 1994, "Evidence that Volume Sustains Price Changes, *Financial Analysts Journal* (November-December), 57-67.

Wang, Jiang, 1994, "A Model of Competitive Stock Trading Volume", *Journal of Political Economy* 102, 127-168.

Wermers, Russ, 1998, "Mutual Fund Herding and the Impact on Stock Prices", *Journal of Finance.*

Table 1. Descriptive Statistics

	All	Small firms	Large firms	1st subperiod	2nd subperiod
No. of firms	1581	790	791	1483	1679
Market Cap ($ Million)	3850	358	7336	2797	4906
Price	31.39	20.44	42.12	29.65	33.13
Turnover (%)	0.309	0.280	0.335	0.278	0.340
Average FLOW (%)	0.005	-0.001	0.009	0.003	0.007
Std Dev FLOW (%)	0.094	0.104	0.084	0.092	0.096
Std Dev RET (%)	1.646	1.770	1.495	1.473	1.819
Corr of FLOW and RET (%)	32.9	41.3	34.2	34.4	31.4

Table reports the time-series average of daily cross-sectional summary statistics. Sample is NYSE common stocks from 1993 to 1999, excluding stocks priced under five dollars.

Turnover : daily shares traded, scaled by shares outstanding.

FLOW : the estimated daily sum of buyer-initiated shares minus seller-initiated shares traded, scaled by total shares outstanding.

RET : daily stock return, computed from closing mid-quotes (adjusted for any stock splits or dividends).

Table 2. Bivariate Regressions of Returns and Flows

Equation		(1)				(2)			
Dependent Variable:		RET_t				$FLOW_t$			
Average R-squared		0.103				0.107			
Lagged Regressors:		RET_{t-j}		$FLOW_{t-j}$		RET_{t-j}		$FLOW_{t-j}$	
F-test		[< 0.001]		[< 0.001]		[< 0.001]		[< 0.001]	
Lagged	lag 1	0.0031	(2.19)	0.0591	(3.85)	0.0023	(43.80)	0.0854	(77.03)
Coefficients	lag 2	-0.0191	(-15.24)	-0.0874	(-6.06)	0.0004	(9.63)	0.0461	(45.29)
	lag 3	-0.0106	(-9.53)	-0.0607	(-4.38)	0.0000	(1.42)	0.0336	(34.72)
	lag 4	-0.0111	(-10.27)	0.0045	(0.33)	-0.0001	(-2.40)	0.0258	(27.45)
	lag 5	-0.0093	(-9.01)	0.0133	(0.94)	-0.0002	(-5.44)	0.0222	(22.26)
	lags 6-40				omitted for brevity				
Cumulative	5-day	-0.0468	[< 0.001]	-0.0814	[0.016]	0.0030	[< 0.001]	0.2468	[< 0.001]
Impulse	10-day	-0.0615	[< 0.001]	-0.1672	[< 0.001]	0.0015	[< 0.001]	0.3706	[< 0.001]
Response	20-day	-0.0703	[< 0.001]	-0.3789	[< 0.001]	-0.0016	[< 0.001]	0.5602	[< 0.001]
Function	40-day	-0.0629	[< 0.001]	-0.8097	[< 0.001]	-0.0074	[< 0.001]	0.8938	[< 0.001]

Table summarizes coefficients estimated from daily cross-sectional regressions of the following two equations, with the number of lags (J) set to 40 days:

$$RET_{i,t} - \overline{RET_i} = \hat{r_t} + \sum_{j=1}^{j=J} \hat{a_j}(RET_{i,t-j} - \overline{RET_i}) + \sum_{j=1}^{j=J} \hat{b_j}(FLOW_{i,t-j} - \overline{FLOW_i}) + \hat{u_{i,t}} \quad (1)$$

$$FLOW_{i,t} - \overline{FLOW_i} = \hat{f_t} + \sum_{j=1}^{j=J} \hat{c_j}(RET_{i,t-j} - \overline{RET_i}) + \sum_{j=1}^{j=J} \hat{d_j}(FLOW_{i,t-j} - \overline{FLOW_i}) + \hat{v_{i,t}} \quad (1)$$

Mean coefficients and standard errors are derived from the time series of the estimates. R-squareds are the average of the daily cross-sectional regressions. Cumulative impulse response coefficients, for 5-, 10-, 20- and 40-days, are computed from the estimated coefficients as described in section 3 (p-values shown in squared brackets are derived via Monte Carlo simulation). F-tests of the joint coefficient significance of each of the four blocks of regressors are derived from the time-series mean and standard errors of the coefficient estimates (p-values are shown in squared brackets). Sample is NYSE common stocks from 1993 to 1999, excluding stocks priced under five dollars.

Table 3. Bivariate regressions in subperiods

Equation		(1)				(2)			
Dependent Variable:		RET_t				$FLOW_t$			
Lagged Regressors:		RET_{t-j}		$FLOW_{t-j}$		RET_{t-j}		$FLOW_{t-j}$	
Panel A. First Half Subperiod, All stocks									
Average R-squared		0.098				0.108			
	1-day	0.0013	[0.496]	0.1247	[< 0.001]	0.0024	[< 0.001]	0.0785	[< 0.001]
Cumulative	5-day	-0.0609	[< 0.001]	0.0643	[0.114]	0.0028	[< 0.001]	0.2243	[< 0.001]
Impulse	10-day	-0.0777	[< 0.001]	-0.0233	[0.670]	0.0013	[< 0.001]	0.3365	[< 0.001]
Response	20-day	-0.0818	[< 0.001]	-0.1872	[0.026]	-0.0017	[< 0.001]	0.5188	[< 0.001]
Function	40-day	-0.0744	[< 0.001]	-0.7075	[< 0.001]	-0.0079	[< 0.001]	0.8556	[< 0.001]
Panel B. Second Half Subperiod, All stocks									
Average R-squared		0.108				0.105			
	1-day	0.0051	[0.019]	-0.0063	[0.780]	0.0023	[< 0.001]	0.0923	[< 0.001]
Cumulative	5-day	-0.0324	[< 0.001]	-0.2341	[< 0.001]	0.0031	[< 0.001]	0.2698	[< 0.001]
Impulse	10-day	-0.0447	[< 0.001]	-0.3236	[< 0.001]	0.0016	[< 0.001]	0.4059	[< 0.001]
Response	20-day	-0.0585	[< 0.001]	-0.5857	[< 0.001]	-0.0015	[< 0.001]	0.6038	[< 0.001]
Function	40-day	-0.0512	[< 0.001]	-0.9182	[< 0.001]	-0.0069	[< 0.001]	0.9334	[< 0.001]
Panel C. Full Period excluding Dec, Jan, Feb months, All stocks									
Average R-squared		0.103				0.107			
	1-day	0.0024	[0.145]	0.0680	[< 0.001]	0.0024	[< 0.001]	0.0867	[< 0.001]
Cumulative	5-day	-0.0455	[< 0.001]	-0.0799	[0.030]	0.0030	[< 0.001]	0.2498	[< 0.001]
Impulse	10-day	-0.0583	[< 0.001]	-0.1743	[< 0.001]	0.0014	[< 0.001]	0.3745	[< 0.001]
Response	20-day	-0.0641	[< 0.001]	-0.4420	[< 0.001]	-0.0019	[< 0.001]	0.5581	[< 0.001]
Function	40-day	-0.0545	[< 0.001]	-0.9469	[< 0.001]	-0.0079	[< 0.001]	0.8841	[< 0.001]

Table summarizes coefficients estimated from daily cross-sectional regressions of the following two equations, with the number of lags (J) set to 40 days:

$$RET_{i,t} - \overline{RET_i} = \hat{r} + \sum_{j=1}^{j=J} \hat{a}_j(RET_{i,t-j} - \overline{RET_i}) + \sum_{j=1}^{j=J} \hat{b}_j(FLOW_{i,t-j} - \overline{FLOW_i}) + \hat{u_{i,t}} \quad (1)$$

$$FLOW_{i,t} - \overline{FLOW_i} = \hat{f} + \sum_{j=1}^{j=J} \hat{c}_j(RET_{i,t-j} - \overline{RET_i}) + \sum_{j=1}^{j=J} \hat{d}_j(FLOW_{i,t-j} - \overline{FLOW_i}) + \hat{v_{i,t}} \quad (2)$$

Mean coefficients and standard errors are derived from the time series of the cross-sectional estimates, as described in Table 2. Sample is NYSE common stocks from 1993 to 1999, excluding stocks priced under five dollars.

Table 4. Bivariate regressions in size-sorted subsamples

Equation		(1)				(2)			
Dependent Variable:		RET_t				$FLOW_t$			
Lagged Regressors:		RET_{t-j}		$FLOW_{t-j}$		RET_{t-j}		$FLOW_{t-j}$	
Panel A. Large stocks									
Average R-squared		0.191				0.183			
	1-day	-0.0229	[< 0.001]	0.1474	[< 0.001]	0.0037	[< 0.001]	0.0895	[< 0.001]
Cumulative	5-day	-0.0859	[< 0.001]	0.0064	[0.889]	0.0061	[< 0.001]	0.2566	[< 0.001]
Impulse	10-day	-0.0974	[< 0.001]	-0.1698	[0.010]	0.0047	[< 0.001]	0.3774	[< 0.001]
Response	20-day	-0.1049	[< 0.001]	-0.4311	[< 0.001]	0.0013	[0.003]	0.5568	[< 0.001]
Function	40-day	-0.1063	[< 0.001]	-0.7181	[< 0.001]	-0.0043	[< 0.001]	0.8748	[< 0.001]
Panel B. Small stocks									
Average R-squared		0.142				0.168			
	1-day	0.0209	[< 0.001]	0.0021	[0.920]	0.0013	[< 0.001]	0.0740	[< 0.001]
Cumulative	5-day	-0.0180	[< 0.001]	-0.1326	[0.003]	0.0008	[< 0.001]	0.2226	[< 0.001]
Impulse	10-day	-0.0349	[< 0.001]	-0.1725	[0.005]	-0.0005	[0.036]	0.3419	[< 0.001]
Response	20-day	-0.0460	[< 0.001]	-0.3540	[< 0.001]	-0.0033	[< 0.001]	0.5310	[< 0.001]
Function	40-day	-0.0331	[< 0.001]	-0.8431	[< 0.001]	-0.0085	[< 0.001]	0.8491	[< 0.001]

Table summarizes coefficients estimated from daily cross-sectional regressions of the following two equations, with the number of lags (J) set to 40 days:

$$RET_{i,t} - \overline{RET_i} = \hat{r} + \sum_{j=1}^{j=J} \hat{a}_j (RET_{i,t-j} - \overline{RET_i}) + \sum_{j=1}^{j=J} \hat{b}_j (FLOW_{i,t-j} - \overline{FLOW_i}) + \hat{u_{i,t}} \quad (1)$$

$$FLOW_{i,t} - \overline{FLOW_i} = \hat{f} + \sum_{j=1}^{j=J} \hat{c}_j (RET_{i,t-j} - \overline{RET_i}) + \sum_{j=1}^{j=J} \hat{d}_j (FLOW_{i,t-j} - \overline{FLOW_i}) + \hat{v_{i,t}} \quad (2)$$

Mean coefficients and standard errors are derived from the time series of the cross-sectional estimates, as described in Table 2. Sample is NYSE common stocks from 1993 to 1999, excluding stocks priced under five dollars.

Table 5. Univariate Regressions

Equation		(1A)		(1B)		(2A)		(2B)	
Dependent Variable:		RET_t		RET_t		$FLOW_t$		$FLOW_t$	
Lagged Regressors:		RET_{t-j}		$FLOW_{t-j}$		RET_{t-j}		$FLOW_{t-j}$	
Average R-squared		0.072		0.040		0.040		0.073	
Lagged	lag 1	0.0056	(3.97)	0.1157	(6.45)	0.0041	(73.86)	0.1029	(95.21)
Coefficients	lag 2	-0.0205	(-17.04)	-0.2223	(-14.32)	0.0017	(36.55)	0.0487	(50.30)
	lag 3	-0.0118	(-11.05)	-0.1372	(-9.51)	0.0011	(24.80)	0.0329	(36.36)
	lag 4	-0.0114	(-10.85)	-0.0656	(-4.35)	0.0007	(17.03)	0.0242	(27.20)
	lag 5	-0.0089	(-8.99)	-0.0391	(-2.64)	0.0005	(12.85)	0.0202	(21.47)
	lags 6-40			omitted for brevity					
Sum of	5 lags	-0.0470	[< 0.001]	-0.3487	[< 0.001]	0.0083	[< 0.001]	0.2291	[< 0.001]
Lagged	10 lags	-0.0660	[< 0.001]	-0.4675	[< 0.001]	0.0095	[< 0.001]	0.2985	[< 0.001]
Coefficients	20 lags	-0.0821	[< 0.001]	-0.6031	[< 0.001]	0.0104	[< 0.001]	0.3885	[< 0.001]
	40 lags	-0.0867	[< 0.001]	-0.7762	[< 0.001]	0.0116	[< 0.001]	0.5186	[< 0.001]

Table summarizes coefficients estimated from daily cross-sectional regressions of the following four equations, with the number of lags (J) set to 40 days:

$$RET_{i,t} - \overline{RET_i} = \hat{a_0} + \sum_{j=1}^{j=J} \hat{a_j}(RET_{i,t-j} - \overline{RET_i}) + \hat{w_{i,t}} \quad (1A)$$

$$RET_{i,t} - \overline{RET_i} = \hat{b_0} + \sum_{j=1}^{j=J} \hat{b_j}(FLOW_{i,t-j} - \overline{FLOW_i}) + \hat{x_{i,t}} \quad (1B)$$

$$FLOW_{i,t} - \overline{FLOW_i} = \hat{c_0} + \sum_{j=1}^{j=J} \hat{c_j}(RET_{i,t-j} - \overline{RET_i}) + \hat{y_{i,t}} \quad (2A)$$

$$FLOW_{i,t} - \overline{FLOW_i} = \hat{d_0} + \sum_{j=1}^{j=J} \hat{d_j}(FLOW_{i,t-j} - \overline{FLOW_i}) + \hat{z_{i,t}} \quad (2B)$$

Mean coefficients and standard errors are derived from the time series of the cross-sectional estimates. R-squareds are the average of the daily regressions. Sample is NYSE common stocks from 1993 to 1999, excluding stocks priced under five dollars.

Table 6. Bivariate Time Series Regressions

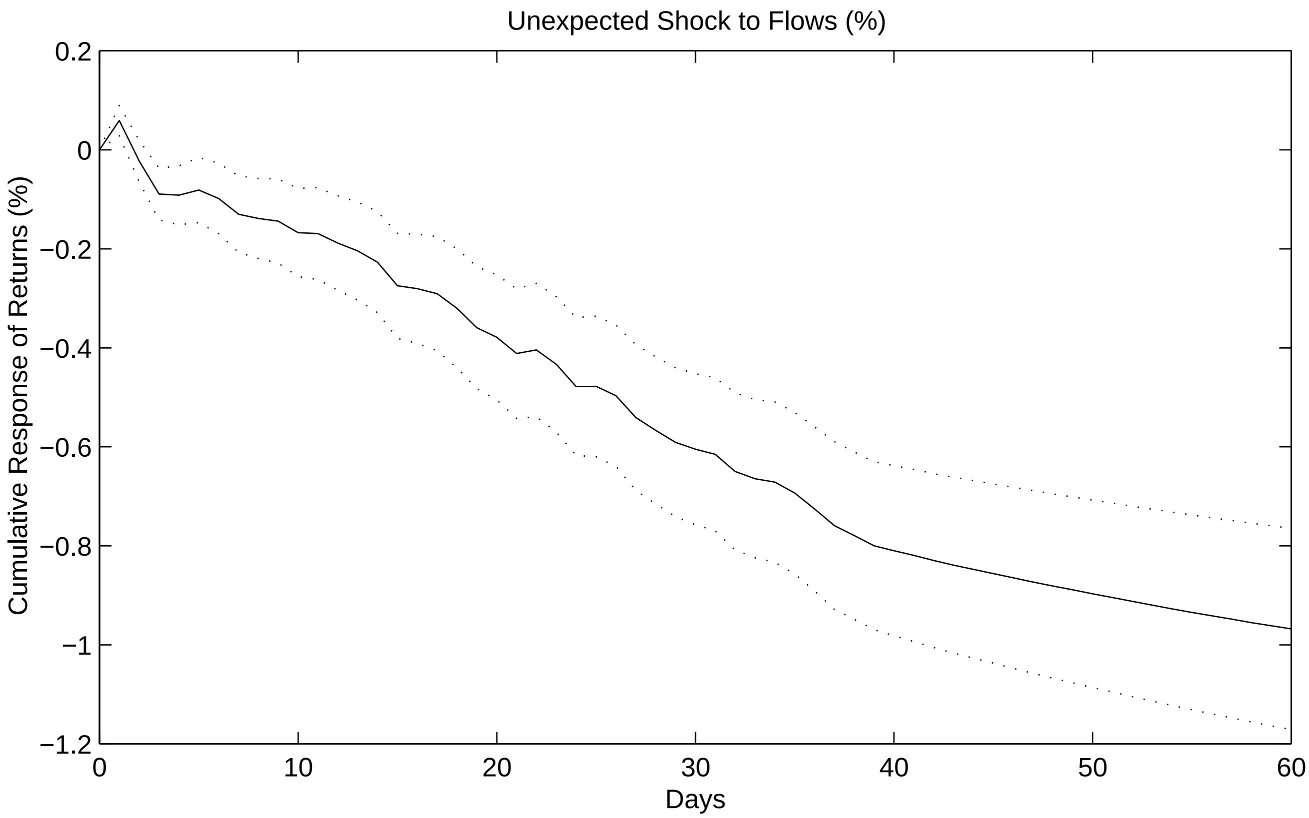
Equation		(1)				(2)			
Dependent Variable:		RET_t				$FLOW_t$			
Average R-squared		0.103				0.107			
Lagged Regressors:		RET_{t-j}		$FLOW_{t-j}$		RET_{t-j}		$FLOW_{t-j}$	
Lagged	lag 1	0.0153	(59.38)	0.0783	(54.97)	0.0025	(84.79)	0.0781	(97.49)
Coefficients	lag 2	-0.0172	(29.65)	-0.0492	(47.10)	0.0007	(65.51)	0.0403	(87.12)
	lag 3	-0.0059	(43.56)	-0.0844	(44.68)	0.0003	(56.18)	0.0294	(80.64)
	lag 4	-0.0111	(35.09)	0.0397	(51.94)	0.0001	(50.22)	0.0221	(73.64)
	lag 5	-0.0083	(39.33)	-0.0354	(48.49)	-0.0000	(46.41)	0.0182	(72.17)
	lags 6-40				omitted for brevity				
Cumulative	5-day	-0.0202	(40.7)	-0.0618	(49.2)	0.0043	(74.4)	0.2201	(99.0)
Impulse	10-day	-0.0189	(43.6)	-0.3041	(45.7)	0.0038	(64.1)	0.3119	(98.4)
Response	20-day	-0.0188	(43.2)	-0.6631	(40.1)	0.0024	(55.7)	0.4424	(97.9)
Function	40-day	-0.0168	(45.1)	-1.3409	(36.6)	-0.0003	(50.3)	0.6246	(96.9)

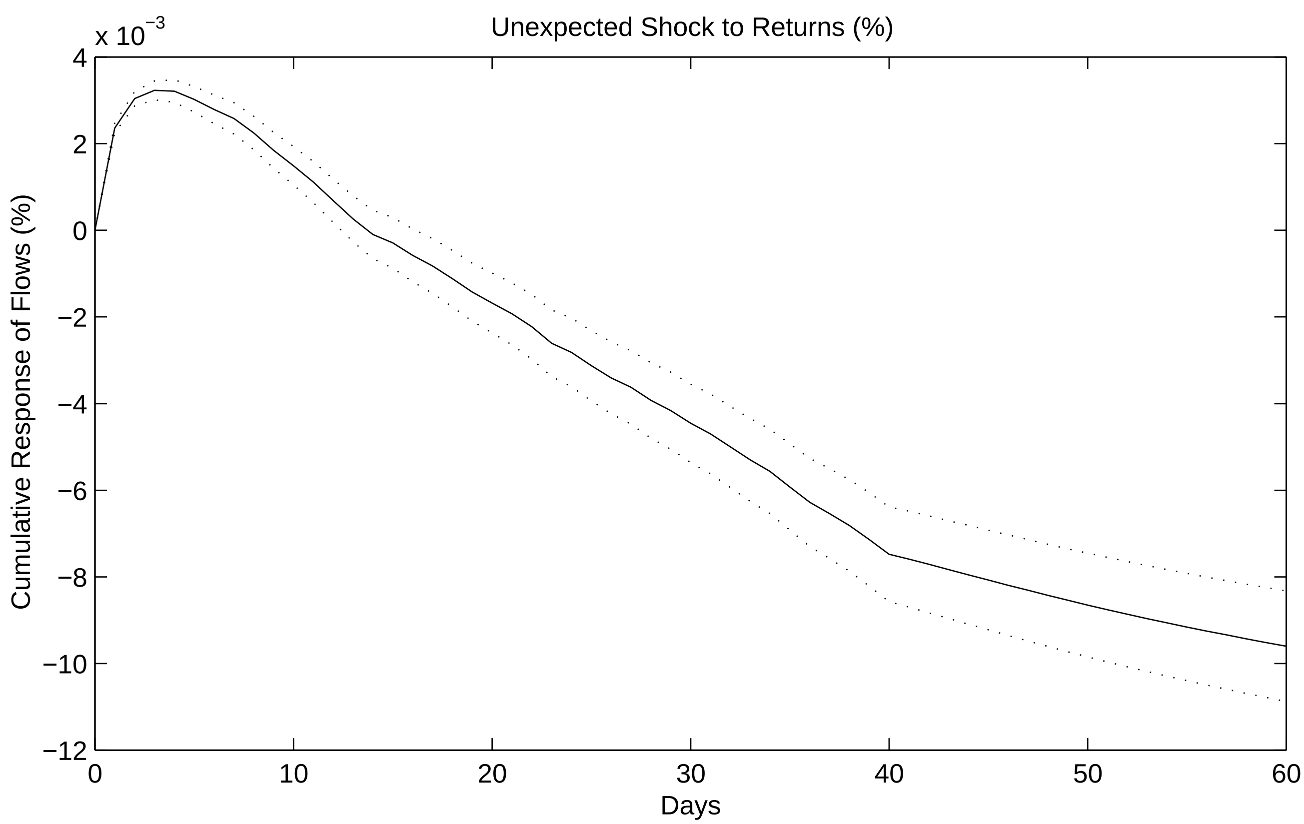
Table summarizes the mean coefficients (percent positive in parentheses), across stocks, estimated from individual time-series regressions of the following two equations, with the number of lags (J) set to 40 days:

$$RET_{i,t} = \hat{r_i} + \sum_{j=1}^{j=J} \hat{a_{i,j}} RET_{i,t-j} + \sum_{j=1}^{j=J} \hat{b_{i,j}} FLOW_{i,t-j} + \hat{u_{i,t}} \quad (1)$$

$$FLOW_{i,t} = \hat{f_i} + \sum_{j=1}^{j=J} \hat{c_{i,j}} RET_{i,t-j} + \sum_{j=1}^{j=J} \hat{d_{i,j}} FLOW_{i,t-j} + \hat{v_{i,t}} \quad (2)$$

Cumulative impulse response coefficients, for 5-, 10-, 20- and 40-days, are computed from the estimated coefficients as described in section 3 (percent positive in parentheses). Sample is 1157 NYSE common stocks from 1993 to 1999, with at least 5 years of data available, priced over five dollars.


Unexpected Shock to Flows (%)
Cumulative Response of Returns (%)
Days
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−1.2
0
10
20
30
40
50
60


Unexpected Shock to Returns (%)
x 10^−3
Cumulative Response of Flows (%)
Days
4
2
0
−2
−4
−6
−8
−10
−12
0
10
20
30
40
50
60