Buyers Versus Sellers: Who Initiates Trades And When?

Tarun Chordia, Amit Goyal, and Narasimhan Jegadeesh*

September 2012

Abstract

We examine the relation between the flow of orders from buyers and sellers, and past return and stock characteristics. We find that the difference between buyer- and seller-initiated trades (the order imbalance) is negatively related to short horizon returns, but positively related to returns over longer horizons. We also find strong evidence of tax-motivated trading around the turn of the year. We do not find support for the flight-to-quality hypothesis, which posits selling as market risk increases.

Contacts

	Chordia	Goyal	Jegadeesh
Voice:	1-404-727-1620	+41 21 692 3676	1-404-727-4821
Fax:	1-404-727-5238	+41 21 692 3435	1-404-727-5238
E-mail:	Tarun_Chordia@bus.emory.edu	amit.goyal@unil.ch	Narasimhan_Jegadeesh@bus.emory.edu
Address:	Goizueta Business School Emory University Atlanta, GA 30322	Swiss Finance Institute at the University of Lausanne 1015 Lausanne, Switzerland	Goizueta Business School Emory University Atlanta, GA 30322

* We thank Edward Dyl, Jean Helwege, Eric Kelley, Christopher Lamoureux, Maureen O'Hara, finance doctoral students at Emory University and National University of Singapore, and seminar participants at Exeter University, University of Arizona, University of New South Wales, University of Queensland, University of South Carolina, University of Sydney, University of Technology Sydney and Vienna University of Economics and Business for helpful comments. Amit Goyal would like to thank Rajna Gibson for her support through her NCCR-FINRISK project.

Buyers Versus Sellers: Who Initiates Trades and When?

September 2012

Abstract

We examine the relation between the order flows from buyers and sellers, and past returns and stock characteristics. We find that the difference between buyer- and seller-initiated trades (order imbalance) is negatively related to short horizon returns, but positively related to returns over longer horizons. We also find strong evidence of tax-motivated trading around the turn of the year. We do not find support for the flight-to-quality hypothesis, which posits that investors are more likely to initiate selling when market risk increases.

Key Words: Order imbalances, Disposition effect, Tax-loss selling, Momentum trading, contrarian trading

Portfolio managers and market makers are keenly interested in understanding how order flows from buyers and sellers are affected by observable factors such as past returns and stock characteristics. Understanding such factors is critically important for trading decisions and price discovery in financial markets. For example, in many market microstructure models market makers infer value-relevant information from buy and sell order flows. If buy and sell orders are systematically driven by past returns, then market makers should correctly update their conditional expectations of order flow to efficiently extract information from these orders. Similarly, portfolio managers likely face less adverse price impacts if they time their orders to trade against expected order flow. In general, a good understanding of factors that affect the direction of order flow provides important insights into time variation in trading volume and expected price impacts of buy and sell trades.

A number of theories in the literature hypothesize that past returns are important determinants of order flows. For example, the disposition effect predicts that investors are more likely to sell winners and predicts a negative relation between past returns and order flows. Other hypotheses such as the momentum trading hypotheses predict an opposite relation between past returns and order flows. Although these hypotheses differ on the sign of the relation, they both predict that order flows would be related to past returns. Tax-induced trading hypothesis and contrarian trading hypotheses also predict that order flows would be related to past returns.

Our empirical analysis examines the relation between order flows from buyers and sellers, and past returns. We test the validity of the underlying hypotheses based on the sign and significance of these relations. Several studies in the literature examine some of these theories using data on trades of individual investors and institutions. While these studies focus on the actions of individual investors, our focus is on aggregate order flows. We address the larger issue of whether the actions of individual investors, as predicted by various hypotheses, are pervasive enough to impact aggregate trading behavior.

At the aggregate level, several studies in the literature use turnover data to test some of the hypotheses that we examine. In contrast, we use measures of buyer- and seller-initiated trades in our tests since the underlying hypotheses make predictions about when buyers or sellers would initiate trades and not about turnover per se. Without knowing whether a buyer or seller initiated the trade, it is hard to interpret any relation between trading volume/turnover and past returns as being consistent with any particular theory. Moreover, understanding the factors that affect the direction of trade provides a different perspective that is particularly useful for market makers and traders who seek to extract information from order flow and to strategically time buy and sell trades.

Briefly, the predictions about buyer- and seller-initiated trades that we test, and the hypotheses that underlie the predictions are the following:

1. Disposition effect: Sellers are more likely to initiate trades for past winners (stocks with positive past returns) than for past losers (stocks with negative past returns).
2. Tax-induced trading: Sellers are more likely to initiate trades for losers than for winners.
3. Seasonal tax-induced trading: Sellers are more likely initiate trades for losers than for winners in December but more likely to initiate trades for winners in January.
4. Momentum trading: Buyers are more likely to initiate trades for winners and sellers are more likely to initiate trades for losers.
5. Contrarian trading: Buyers are more likely to initiate trades for losers and sellers are more likely to initiate trades for winners.
6. Flight-to-quality: Sellers are more likely to initiate trades when market risk increases and buyers are more likely to initiate trades when market risk decreases.

We find the following evidence. In time-series tests, order imbalances between buyer- and seller-initiated trades are negatively related to one-month lagged returns, but unrelated to returns at longer lags. This negative relation of order imbalance with one-month lagged returns is consistent with either a contrarian trading strategy or with the disposition effect. In any case, this result indicates that the direction of market-wide trades is not related to returns over horizons longer than that of one month. In addition, we do not find any support for the flight-to-quality hypothesis.

In cross-sectional tests, we find that order imbalances are negatively related to returns over the past three months but positively related to returns at longer horizons. These results indicate that the relation between order imbalances and past returns cannot be fully captured by any single hypothesis. The evidence that order imbalances are negatively related to shorter horizon returns supports both the contrarian trading behavior and the disposition effect which both make the same predictions. However, the negative relation between order imbalances and past returns applies to a much shorter time period than investors' typical holding periods, which suggests that the aggregate effect is more likely to be driven by short horizon contrarian traders.

We find strong support for the seasonal tax-trading hypothesis. In December, investors are more likely to sell losers in order to realize capital losses, but less likely to sell winners. In contrast, investors are more likely to sell winners in January, which enables them to defer realization of capital gains to the following year. In addition to past returns, we find that market capitalization, book to market ratio and past order imbalances are useful in predicting future order imbalances.

The rest of the paper is organized as follows. The next section discusses the hypotheses and briefly reviews the literature. Section II discusses order imbalances and presents the data. Section III discusses the results. Robustness tests are presented in Section IV and Section V concludes.

I. Determinants of Buy and Sell Order Flows

Several hypotheses proposed in the literature predict that order flow is related to past returns. The disposition effect posits that investors are likely to sell their winners but are reluctant to sell their losers. Shefrin and Statman (1985) develop this concept based on prospect theory proposed by Kahneman and Tversky (1979). The prospect theory posits that individuals' utility function is concave for gains and convex for losses; Shefrin and Statman argue that such utility functions lead to the disposition effect. Therefore, the disposition effect predicts that sellers are more likely to initiate trades for winners than for losers.

Constantinides' (1984) tax-motivated trading model makes the opposite prediction. Constantinides shows that, if there were no trading costs or limits on capital loss deductions, investors should optimally sell all stocks that experience capital losses immediately, and that they should not sell their winners. If investors follow this policy, they are more likely to initiate trades for losers than for winners. However, because of trading costs and limits on capital loss deductions, investors would not continuously follow such a strategy, and any tax-induced trading may be concentrated around the turn of the year (for example, see Lakonishok and Smidt, 1986 and Poterba and Weisbenner, 2001). Under this seasonal tax-loss selling hypothesis, investors are more likely to sell past losers than winners in December to realize capital losses, and are more likely to defer realization of capital gains and sell winners in January.

The hypotheses discussed thus far focus on the motives of sellers. When do buyers initiate trades? Several models predict that uniformed investors would rationally follow trend chasing or momentum strategies when they trade. For example, in Wang (1993), Brennan and Cao (1997), and Hong and Stein (1999), uninformed investors extract informed investors' signals from price changes and trade in the same direction as past returns. Therefore, under the momentum trading hypothesis, we expect more buyer-initiated trades for past winners and more seller-initiated trades for past losers.

In addition to the cross-sectional implications of the theories we discuss above, some of the theories also have implications for buyer- and seller-initiated trades at the aggregate level. For example, since there are likely to be more winners than losers following up-markets than down-markets, the disposition effect predicts that there would be more aggregate seller-initiated trades following up markets. The model of Orosel (1998), however, makes the opposite prediction. Orosel assumes that investors face a fixed cost to participate in the stock market and shows that market participation increases following high market returns and decreases following low market returns. Therefore, this model predicts a positive relation between buyer-initiated trades and past returns.

We also test the popular idea that when the risk or market volatility increases, there is a "flight to quality." The flight-to-quality hypothesis posits that when markets get riskier, investors reduce their stock holdings, and flee to safer assets such as Treasury Bonds.[1] This hypothesis implies that sellers initiate trades when market risk increases, and hence we expect a negative correlation between various measures of market risk and order imbalances.

Several papers use turnover data to test some of the hypotheses we discuss above. However, since these theories make predictions about whether buyers or sellers initiate trades, it is hard to test them with turnover data that do not identify who initiates trades. For example, Statman, Thorley, and Vorkink (2006) find that turnover is positively related to past returns and conclude that this evidence supports the disposition effect.[2] This interpretation assumes, however, that the changes in turnover are driven by seller-initiated trades. The evidence in Statman, Thorley, and Vorkink could well be interpreted as evidence of momentum trading

[1] Caballero and Krishnamurthy (2008) develop a model of flight-to-quality based on liquidity shortages and Knightian uncertainty. They show that an increase in Knightian uncertainty or a decrease in aggregate liquidity can generate flight-to-quality. Beber, Brandt, and Kavajecz (2009) show that in the Euro-area government bond market, credit quality is an important determinant of bond valuation but during periods of market stress, investors chase liquidity, not credit quality.

[2] Odean (1998) uses individual investor stock trading data to provide support for the disposition effect (see also Ben-David and Hirshleifer, 2012). We examine aggregate investor behavior.

rather than evidence of the disposition effect if one were to assume that buyers, rather than sellers, initiate trades for past winners.

Since both disposition effect and momentum trading make the same predictions for turnover, tests with turnover data cannot differentiate between these hypotheses. However, these two hypotheses make opposite predictions for order imbalances. Under the disposition effect, we expect a negative relation between order imbalances and past returns, while we expect a positive relation under the momentum trading hypothesis.

II. Trade Initiation and Order Imbalances

In any trade, there is a buy side and a sell side. Generally, one side actively initiates a trade and the other side passively completes the trade. As many microstructure models characterize, the active side demands immediacy and hence this side is willing to pay the passive side a price concession for providing liquidity to complete the trade. For example, a buyer-initiated trade would typically occur at the ask price, and the difference between the ask price and the mid-point of the bid and ask prices is the compensation to the passive side.

A natural question that arises is why do the theories that we examine apply to active side and not the passive side? We can best answer this question within the context of a simple setting that serves as a basis for many market microstructure models. As in Kyle (1985), Admati and Pfleiderer (1988), and others, we can categorize the market participants as market makers, informed traders, and noise traders. The market makers' role is to provide liquidity to the market; they take the opposite side of the net order flow to clear the market. The theories that we examine do not apply to market makers because they do not initiate trades but passively ensure market clearing.

Informed traders trade based on private information; their motive for trade is to profit from such information. Therefore, the motives proposed in the theories that we test do not apply to informed trades. By definition, any private information that these informed traders use is orthogonal to any public information. Our tests, on the other hand, examine the relation between order flow and information in the public domain, such as past returns and volatility. Thus, even though our measure of order flow includes trades initiated by informed traders, these trades contribute to noise in our measures but they do not lead to any bias in the relation that we may find between order flow and the variables we use (more formally, informed trades form part of the error term in a regression of order flow on public information variables).

Consider now the investors broadly labeled as noise traders. Their motives are typically outside the scope of microstructure models, but the theories that we examine posit various hypotheses to explain why they trade. For example, an investor who initiates a trade because of the disposition effect may sell stock from her portfolio that has appreciated in value, but market microstructure models are agnostic about her motives and classify her as a noise trader. In this setting, the market maker posts bid and ask quotes and the noise traders initiate trades. Thus, these noise traders are the active trade initiators who underlie our hypotheses.

In an idealized setting where the noise traders/trade initiators place a market order that executes against a limit order or against posted quotes, we can accurately identify the side that initiated the trade based on whether it was executed at the ask price or the bid price. Although such a setting is not a complete depiction of real world stock markets, the literature typically uses such an algorithm to identify the side that initiates the trade because trade initiators are more likely to place market orders since they are less likely to accept the risk of not consummating their trades. In contrast, market makers are more likely to place limit orders because their primary objective is not to take a position in the stock but to clear the market and to profit from the spread.

It is possible that some noise traders/trade initiators may place limit orders rather than market orders with the hope of executing trades at a more favorable price. In such instances, trades would cross between two trade initiators rather than between an active trade initiator and a market maker. Such trades add noise to the measure of order flow but as long as some of the trades are truly between active trade initiators and market makers, the net order flow is still an unbiased measure of net trades initiated for a particular stock. The key point to note here is that market makers are involved in at least some of the trades as liquidity providers, and the hypotheses that we examine do not apply to them.

Of course, if a large fraction of active trade initiators execute their trades through limit orders then the noise in our measure would overwhelm the signal. This would make it difficult to find any relation between order imbalances and past returns, and detect the underlying motivation for trade. The use of limit orders by trade initiators has become more prevalent in the post-decimalization period partly because of narrowing spreads. Additionally, commercial trade execution tools that allow for easier execution through limit orders have become more widely available in recent periods. Since these practices are more prevalent during the post-decimalization period, we expect to find a weaker relation in the more recent part of our sample period than in the earlier part.

In contrast to the more recent period, the Nasdaq market in the pre-1997 period provides an ideal setting for our purposes. During that period, investors traded with market makers at quoted bid and ask prices. Therefore, order flows in pre-1997 Nasdaq data are likely to be less noisy and more informative about the underlying motives for trade. We accordingly examine the robustness of our results using pre-1997 Nasdaq data in greater detail in Section IV.C.

A. Data

Our tests examine the determinants of order imbalances (OIB), which we construct using Trade and Quote (TAQ) data. In addition to TAQ data, we also use data from CRSP and COMPUSTAT. Our sample comprises common stocks listed on the NYSE in the January 1993 through December 2010 period. The sample is comprised of all stocks that satisfy the following criteria in any given month: (i) The return in the current month and over at least the past twelve months are available from CRSP, (ii) sufficient data are available to calculate market capitalization at the end of the previous month, (iii) book value data are available from COMPUSTAT as of the previous calendar year end,[3] (iv) intra-day transactions data over the previous month are available on TAQ. We include only common stocks with share codes 10 or 11 on CRSP. The last criterion filters out ADRs, units, American Trust components, closed-end funds, preferred stocks and REITs. There are 1,521 stocks per month on average that satisfy all the above criteria.

The TAQ dataset reports intraday quotes and prices and quantity of each trade. We use the filtering rules in Chordia, Roll, and Subrahmanyam (2001) to eliminate obvious data recording errors in the TAQ dataset. We then use the Lee and Ready (1991) algorithm to classify transactions as either a buy or a sell. Briefly, we implement the Lee and Ready algorithm as follows: if a trade is executed at a price above (below) the quote midpoint, we classify it as a buy

[3] Book values are calculated as in Fama and French (1992). We define book-to-market ratio at the end of a month as the ratio of the most recently available book value to the market capitalization at the end of the same month. We assume that book values are publicly available six months after fiscal year-end. Stocks with negative book values are not included in the sample in the year in which their book values are negative.

(sell); if a trade occurs exactly at the quote mid-point, we sign it using the previous transaction price according to the tick test (i.e., a buy if the sign of the last nonzero price change is positive and vice versa). The Lee and Ready algorithm uses the fact that seller-initiated trades tend to execute at a lower price than buyer-initiated trades.

We apply the tick test up to the past five price changes. If the past five price changes are zero then we do not use it in the computation of order imbalances. As Lee and Ready (1991) note, the timestamps on quotes are not always correctly synchronized with those for trades and hence they recommend that the quotes be matched to trades with a five-second delay. We follow this five-second delay rule until 1998. Since such recording errors are not observed in the more recent data (see, for example, Madhavan et al., 2002 as well as Chordia, Roll, and Subrahmanayam, 2005) we do not impose any delays after 1998.

We compute the following two measures of OIBs for each month based on number of trades and the number of shares traded:

$$\begin{aligned}
\text{OIBNUM} &= \frac{\text{Number of buyer-initiated trades} - \text{Number of seller-initiated trades}}{\text{Total number of trades}} \\
\text{OIBSH} &= \frac{\text{Number of buyer-initiated shares} - \text{Number of seller-initiated shares}}{\text{Total number of shares traded}}.
\end{aligned} \tag{1}$$

One concern with the Lee and Ready (1991) algorithm is that it may misclassify the side that initiated a particular trade, even if the trade initiator places a market order. Lee and Radhakirshna (2000) and Odders-White (2000) examine the trade-level accuracy of the Lee and Ready algorithm for NYSE traded stocks and report accuracy rates of 93% and 85%, respectively. Both Lee and Radhakirshna and Odders-White use data from the pre-decimalization era, and it is important to assess the reliability of the Lee and Ready algorithm in the post-decimalization era as well. The most recent study that examines this issue is Bidisha, Moulton, and Shkilko (2012). They find that the transaction level accuracy of the Lee and Ready algorithm during the June to December 2005 period is about 68%. Bidisha, Moulton, and Shkilko, however, is not directly comparable to Lee and Radhakirshna and Odders-White because it examines Nasdaq stocks, and focuses solely on short sales. Ellis, Michaely, and O'Hara (2000) is more directly comparable to Bidisha, Moulton, and Shkilko because the former also examine the pre-decimalization period accuracy of the Lee and Ready algorithm with Nasdaq stocks. Ellis, Michaely, and O'Hara find an accuracy rate of 81%. Although the lower accuracy rate in

Bidisha, Moulton, and Shkilko may be partly due to the fact that it focuses only on short sales, it is quite likely that decimalization contributed to this phenomenon as well.

What is important from the perspective of our study, however, is not the trade-level accuracy, but the accuracy when trade-level classifications are aggregated to obtain monthly trade imbalances. For example, even if a fraction of seller-initiated trades on a particular day is misclassified as buyer-initiated trades and a similar fraction of buyer-initiated trades is also misclassified, then daily-level accuracy would be much greater than trade-level accuracy. In fact, Bidisha, Moulton, and Shkilko (2012) find that daily-level error rate is close to zero, and statistically insignificant. Since the error rate for NYSE stocks were lower than the error rates for Nasdaq stocks in the pre-decimalization period, it is quite likely that the error rates in the post-decimalization period would be no worse for NYSE stocks than those for Nasdaq stocks. Also, note that our main tests use monthly-level aggregation, which would further reduce the impact of any trade-level classification error. Therefore, any trade-level misclassification is unlikely to meaningfully impact our results.

B. Descriptive statistics

Table 1 presents the summary statistics for OIBNUM and OIBSH. OIBNUM assigns the same weight to a trade regardless of trade size and hence a small trade will get the same weight as a large trade using this measure. Since individuals typically trade smaller quantities relative to institutions, this measure is weighted more towards small investors as compared to OIBSH. OIBSH weights order imbalances by trade sizes, and hence it gives large institutional trades more weight than small retail trades.

All summary statistics in Table I are time-series averages of the corresponding statistics in the cross-section each month. The mean (median) OIBNUM is 2.5% (3.3%) and the mean (median) OIBSH is 3.0% (4.2%). Therefore, on average, there are more buyer-initiated trades than seller-initiated trades. The market capitalization is positively skewed (sample skewness is 8.8) with mean (median) values of $5.37 ($1.11) billion. The mean (median) book-to-market value is 0.80 (0.54). Table 1 also reports individual stock return volatility, which is the monthly standard deviation. Specifically, for each stock we compute the standard deviation of daily percentage returns each month and multiply it by the square root of the number of trading days during that month. The mean (median) monthly volatility is 12.1% (10.3%).

In order to ascertain whether there are differences in these statistics across different kinds of stocks, we also calculate them for certain subsamples of stocks. Each month, we divide the sample of stocks into small and big stocks based on whether their market capitalization is smaller or larger than the median market capitalization at the beginning of that month. We follow a similar procedure to classify stocks into growth and value stocks based on their book-to-market ratios. Panels B through E of Table 1 present the descriptive statistics for these stocks. We note that the order imbalance is higher for big stocks than that for small stocks, and for growth stocks than that for value stocks. Both OIBNUM and OIBSH show these differences.

Figure 1 presents value-weighted average OIBs over time. Both, OIBNUM and OIBSH are in general positive except for a few months in the early part of the sample period when the OIBNUM is sometimes negative, and in the period starting in the latter half of 2007 when both measures of OIB turn negative. Figure 1 also indicates that OIBNUM is smaller than OIBSH almost over the entire sample period. This finding suggests that, as compared to small investors, institutions are more likely to initiate buys.

Figure 2 presents order imbalances for various quintiles partitioned based on the market capitalizations of the stocks in the sample. In this figure, Small, Mid and Large represents the quintiles of smallest, middle, and largest firms in the sample, Order imbalances are generally smaller for small stocks than those for medium stocks, which are in turn lower than those for large stocks. In fact, order imbalances are generally negative for small stocks over much of the sample period. Since institutional holdings are positively related to firm size, these results also suggest that institutions are more likely to initiate buys than are small investors.

III. Empirical Tests and Results

We examine both the time-series and the cross-sectional determinants of order imbalances. In both our time-series and cross-sectional tests we have to make some choices. Although the hypotheses that form the basis of our tests predict that trading decisions are related to past returns, they do not specify the horizons over which returns affect trading. There is no theory to guide us on the correct number of lagged returns to be used. For instance, is it the returns over the past three months or those over the past six months that are likely to drive the disposition effect or tax-motivated trading? The theories suggest that investors take into account the prices at which they bought the stocks when they decide to trade, but different investors buy stocks at different points in time. In practice, both individuals and institutions have widely varying investment horizons.[4] Therefore, our empirical tests consider past 12 monthly returns to let the data tell us about the impact of returns at various horizons on investors' trading decisions.[5]

A. Time-series determinants of order imbalances

We first examine the determinants of order imbalances in the time-series. Specifically, we form value-weighted, size sorted quintile portfolios[6] and regress order imbalances on past twelve months of returns, past month's portfolio volatility and past OIB as per the following time-series regression:

$$OIB_{i,t} = a_i + \sum_{j=1}^{12} b_{i,j} R_{i,t-j} + c_i Vol_{i,t-1} + \sum_{j=1}^{3} d_{i,j} OIB_{i,t-j} + e_{i,t} \ , \qquad (2)$$

where $R_{i,t}$ is the return on portfolio i in month t, and $Vol_{i,t\text{-}1}$ is the standard deviation of the portfolio return computed using daily returns in month $t-1$. We use both OIBNUM and OIBSH as measures of order imbalances in separate regressions. We include lags of OIB because of the high persistence exhibited by this variable. The results we report use three lags of OIB, but we find in unreported tests that additional lags of OIB do not change any of our inferences.

Table 2 presents the coefficient estimates and the corresponding t-statistics for regression (2), which we compute using Newey-West standard errors with three lags. Panels A and B report the results for order imbalances measured in number of trades and measured in terms of number of shares traded, respectively. The adjusted-R^2s of the regressions are fairly high, which is not surprising given the significant coefficients on the lags of OIB.

Focusing first on the results for the portfolio of all stocks, we find that the one-month lagged return coefficient is significantly negative when we use OIBNUM as the independent variable and the one- and two-month lagged return coefficients are significantly negative with OIBSH as the independent variable. Given the monthly volatility of returns of around 4.2%, the coefficient estimates imply a decline in OIBNUM of 0.29% and OIBSH of 0.38% following a one standard deviation positive return in the previous month (the standard deviation of OIBNUM

[4] For example, Coval, Hirshleifer, and Shumway (2002) find that the holding period of clients of a discount brokerage had a mean holding period 378 days and a standard deviation of 321 days.
[5] We also examine the impact of up to prior 36 monthly returns, but the main effects can be discerned with 12 lags.
[6] We use portfolios because we want to test how OIBs respond in the aggregate to market conditions. In our cross-sectional tests that we present in the next subsection we use individual stocks to test the different hypotheses.

and OIBSH is 2.93% and 3.88%, respectively). The coefficient estimates on higher lags are generally insignificant.[7]

Coefficient on lags greater than one for OIBNUM and lags greater than two for OIBSH are also, in general, insignificant for all size quintile portfolios. Except for the smallest quintile, the coefficient on one-month lagged return for other size quintile portfolios is significantly negative for OIBNUM regressions with magnitude similar to that for the portfolio of all stocks. One- and two-month lagged return coefficients in OIBSH regressions are generally significant only for portfolios of bigger stocks. Thus, for the portfolio of smallest stocks, lagged returns are not significantly related to future order imbalances while the results of portfolio of bigger stocks are similar to those for the portfolio of all stocks.[8]

The most significant coefficient in regression (2) is that on the one-month lagged OIB. Therefore, order imbalances seem to be influenced by factors other than past returns and volatility, and these factors are persistent. Perhaps, order imbalances are affected by general market sentiment that we do not capture. In a modified version of regression (2) that excludes lagged order imbalances, we find in unreported results that the slope coefficients on lagged returns are not significant, although the magnitude of the point estimates for the entire sample is similar to those reported in Table 2. Therefore, the inclusion of lagged order imbalances seems to increase the power of these tests because it reduces noise in these regressions.

These results show that, after controlling for the effects of past order flows, investors are more likely to initiate sell trades following an up-market and less likely to do so following down-market. This result does not support the model of Orosel (1998), which predicts increased stock market participation following market price increases and which, therefore, implies a positive coefficient on lagged returns. This negative relation between OIB and one-month lagged returns could result from a contrarian trading strategy, i.e., selling winners and buying losers. The negative coefficients are also consistent with the disposition effect. However, the fact that there is no strong relation between returns at longer lags and order imbalances is somewhat puzzling since holding periods, particularly for individual investors, are longer than two months. Perhaps, investors' mental accounts that drive the disposition effect in Shefrin and Statman (1985) are

[7] We do find that the coefficient estimate on the fifth lag is significantly positive for both OIBNUM and OIBSH. However, it is perhaps not too surprising that we find one significantly positive coefficient when we consider multiple lags.

[8] We found similar results in two equal subperiods, but the slope coefficients were smaller in the second subperiod.

more strongly influenced by the recent performance of the stocks that they hold rather than by the price at which they bought the stocks. In any event, returns over longer horizons do not affect trading decisions at the aggregate level. It is also puzzling that there is no evidence of disposition effect for small stocks, given that retail investors hold a larger proportion of the smaller stocks.

Griffin, Nardari, and Stulz (2007) find that in the U.S. total turnover is also not related to prior returns during the 1993 to 2003 sample period although they find a positive relation between turnover and past return in less developed markets.[9] They note that this positive relation is consistent with many theories including models that posit behavioral biases and participation costs, and they conclude that "As a result, we find support for many of them." Since we examine order imbalances, we get different predictions from these hypotheses and we find that the time-series relation between trading and past returns is not due to participation costs but they are more likely to be due to either the disposition effect or contrarian trading.

We do not find any relation between order imbalances and volatility, which indicates that seller-initiated trades are as likely in periods of high volatility as they are in periods of low volatility. Therefore, our results do not provide any support for the flight-to-quality hypothesis. Connoly, Stivers, and Sun (2005) suggest that their finding of a positive relation between contemporaneous changes in VIX and bond returns suggest that investors may flee from equity to bonds when volatility increases. However, we do not find any direct relation between order imbalances and risk even when we use VIX as a proxy for market risk in equation (2).

To further investigate the relation between market risk and aggregate order imbalances, in untabulated tests, we tried several other variables as proxies for market risk. Specifically, we used the Pástor and Stambaugh (2003) illiquidity factor, the TED spread, and dummy variables for the LTCM crisis, the Russian banking crisis, and the year 2008 as proxies for market risk. None of these variables are significantly related to order imbalances. Therefore, our evidence does not support the hypothesis that, in the time-series, aggregate investor trading behavior responds to risk or liquidity shocks.

Overall, we find that aggregate market order imbalances are negatively related to past one- and two-month lagged returns, but they are not related to returns at longer horizons or to volatility. These results do not support the idea that investors are more likely to participate in the

[9] Statman, Thorley, and Vorkink (2006) find a positive relation between market turnover and prior returns, but their sample period is from 1962 to 2003. Our sample period is more recent and it includes the 1993 to 2003 period in Griffin, Nardari, and Stulz (2007).

market following an up-market than a down-market as predicted by Orosel (1998) or the flight-to-quality hypothesis. However, these results provide some support for a contrarian trading strategy and possibly for the disposition effect at the aggregate level.

B. Cross-sectional determinants of order imbalances

We now examine the cross-sectional determinants of order imbalances. Specifically, we fit the following cross-sectional regression to examine the cross-sectional relation between order imbalances (OIBNUM and OIBSH) and past returns, lagged volatility, and lags of order imbalances:

$$OIB_{i,t} = a_t + \sum_{j=1}^{12} b_{t,j} R_{i,t-j} + c_t \ln\left(Size_{i,t-1}\right) + d_t \ln\left(BM_{i,t-1}\right) + e_t Vol_{i,t-1} + \sum_{j=1}^{3} f_{t,j} OIB_{i,t-j} + u_{i,t} \ , \quad (3)$$

where $Vol_{i,t-1}$ is the standard deviation of stock i returns, which we compute using daily returns in month t–1, ln($Size_{i,t-1}$) is the logarithm of market capitalization of equity at the end of month t–1, and ln($BM_{i,t-1}$) is the logarithm of book-to-market ratio at the end of month t–1. We include size and book-to-market because of the cross-sectional differences in order imbalances between different kinds of stocks, as evidenced in the descriptive statistics in Table 1.

While we use a number of observable control variables in regression (3), order imbalances are likely to be driven by some factors that are excluded from the regression. For example, if certain investors have private information about pending good news about a firm then they would initiate buy trades leading to positive order imbalances. We do not include a proxy for such informed trading because by its very nature, such private information is unobservable.[10]

As we discussed before, any private information is orthogonal to public information. Since all independent variables in regression (3) are in the public domain, they are orthogonal to any private information. Therefore, order imbalances due to trades motivated by private information will be part of the noise term in the regression, and the omission of private information related variables from the regression does not bias any of our regression estimates.

[10] See Sarkar and Schwartz (2009) for an analysis of information related motives of trading.

Table 3 reports the time-series averages of the monthly slope coefficients and the *t*-statistics computed using Newey-West standard errors with three lags. The slope coefficients on firm size are significantly positive and those on book-to-market ratio are significantly negative for both OIBNUM and OIBSH. This result implies that investors initiate more buy trades for large firms than for small firms, and initiate more buys for growth stocks than value stocks. The absolute value of the coefficient on firm size is higher for OIBSH than that for OIBNUM, while that on book-to-market is higher for OIBNUM than that for OIBSH. Since OIBSH is weighted relatively more towards large investors, these results indicate that large investors initiate more buy trades for large firms than for small firms, and that small investors initiate more buys for growth stocks than value stocks.

The slope coefficients on the first three lagged returns are significantly negative and all the other coefficients on returns at longer lags are positive and significant pointing to either the disposition effect or contrarian trading at short horizons and momentum trading at longer horizons. The negative slope coefficients decline in magnitude and significance with increasing lags. For example, with OIBNUM as the dependent variable, the slope coefficients at the first three lags decline (in absolute terms) from −11.09 to −1.33.

The slope coefficients on lagged order imbalances are significantly positive. Chordia, Roll, and Subrahmanyam (2005) suggest that the positive serial correlation in order imbalances could arise due to herding by investors, or because of splitting of large orders. Another possible explanation for the positive autocorrelation is that when there is substantial imbalance, passive traders who absorb the imbalance may later actively liquidate their positions to reduce their inventory risk. Such activities may underlie short-term autocorrelation in order imbalances because market makers who provide liquidity typically offset inventory imbalances within short periods. The longer-term autocorrelation in order imbalances are likely due to some persistent factors that are not captured by returns and volatility, such as investor sentiment about certain stocks or certain sectors.

The inclusion of lagged OIBs in regression (3) raises an interesting issue. Our results indicate that lagged returns impact order imbalances, and hence OIB_{t-1} would be impacted by returns at longer lags. Therefore, it is possible that some of the impact of lagged returns on OIB$_t$ may be absorbed by lagged OIBs on the right-hand side of the regression. However, as we discussed earlier, lagged returns do not fully explain order imbalances. In addition, we find that returns are contemporaneously correlated with order imbalances.[11] Therefore, exclusion of lagged order imbalances from the regression would potentially lead to an omitted variable bias. In results reported in the Appendix, we estimate regression (3) after excluding lagged order imbalances from the right hand side. The findings indicate that the exclusion of lagged order imbalances from the regression does result in an omitted variables bias, and hence we include them in all our regressions. We also check whether three lags of OIB are sufficient in regression (3). In unreported results, we find that inclusion of 12 lags of OIB does not materially change the results. There is a modest improvement in average adjusted-R^2 from 32.04% to 33.95% for OIBNUM and from 15.87% to 17.98% for OIBSH. The magnitude and statistical significance of coefficients on lagged returns remain similar to those reported in Table 3.

Panels B and C of Table 3 also present the results for subperiods from 1993 to 2001 and from 2002 to 2010, respectively. The results are qualitatively similar in both subperiods; the slope coefficients on lagged returns are significantly negative at shorter horizons but generally positive at longer horizons. The results in the first subperiod are stronger than the second subperiod. As we discussed earlier, the weaker results in the second subperiod are likely due to decimalization and innovations in trade execution strategies.

C. Economic significance

This subsection evaluates the economic significance of the variations in order imbalances that can be predicted based on our model. To do so, we first predict order imbalance for each month using the prediction equation below:

$$OIB_{i,t}^{predicted} = \overline{a} + \sum_{j=1}^{12} \overline{b}_j R_{i,t-j} + \overline{c} \ln\left(Size_{i,t-1}\right) + \overline{d} \ln\left(BM_{i,t-1}\right) + \overline{e} Vol_{i,t-1} + \sum_{j=1}^{3} \overline{f}_j OIB_{i,t-j}, \qquad (4)$$

[11] The average contemporaneous cross-sectional correlation between returns and OIBNUM is 0.135 and that between returns and OIBSH is 0.204.

where bars of top of the coefficients denote average of the time-series coefficients from the corresponding regression (3). For each month, we rank stocks based on $OIB_{i,t}^{predicted}$ and form deciles based on this ranking. We judge the economic significance of predictability of the model based on the differences between OIBs of the extreme deciles ten and one.

We find that average difference between extreme deciles is 18.62% for OIBNUM and 15.94% for OIBSH. Since the average cross-sectional standard deviation of OIBNUM and OIBSH is 10.53% and 13.65%, respectively, these differences indicate significant economic predictability.

D. Flight-to-quality

Volatility is significantly negatively related to OIBNUM, but not significantly related to OIBSH in Table 3. For example, the slope coefficients (*t*-statistics) on volatility are −3.55 (−4.17) and 0.31 (0.39) when we use OIBNUM and OIBSH as dependent variables. These results indicate that there are more seller-initiated trades for high volatility stocks than for low volatility stocks among small investors. Institutional investors, who are more heavily represented in OIBSH, seem to be indifferent to volatility in their trading decisions.

The slope coefficient on volatility in the cross-sectional regression, however, does not directly address the flight-to-quality out of equity market hypothesis because it only examines investors' preference for high volatility versus low volatility stocks at any given point in time. However, we can examine whether investors exhibit a stronger preference for low volatility stocks when the overall stock market risk increases. To examine this hypothesis, we examine the slope coefficients during high risk and low risk sub-periods. Specifically, we examine the slope coefficient on volatility in two sub-periods, one when market volatility in the previous month was above the full-sample median and the other when the market volatility in the previous month was below the full-sample median. In unreported results, we found the slope coefficients to be similar in these two sub-periods (−2.63 in periods of high market volatility and −4.53 in periods of low market volatility, with a statistically insignificant difference of 1.90 for the OIBNUM regressions). These conclusions are not sensitive to the classification of months into high and low volatility as we find similar results based on Pástor and Stambaugh (2003) liquidity-based classification.

To summarize, the cross-sectional tests, similar to the time-series tests, also do not offer any support for within-stock market flight-to-quality hypothesis, i.e., the hypothesis that investors flee to safer stocks when market risk increases. However, the negative coefficient on last month's volatility is consistent with the hypothesis that stocks with higher dispersion of opinion (i.e., higher volatility) amongst investors in the presence of short sale constraints leads to overpricing and subsequent selling in these stocks.

E. Tax-loss selling

The negative slope coefficients on short-term lagged returns are inconsistent with the tax-loss selling hypothesis since this hypothesis predicts that investors are more likely to sell losers than winners. However, it is possible that investors' tax-motivated selling of past losers is concentrated in December and that they defer selling their winners until January. Lakonishok and Smidt (1986) find that turnover for losers increases in December and that for winners increases in January and provide support for the seasonal tax-loss selling hypothesis. The evidence in Lakonishok and Smidt is also consistent with strategic trading by sophisticated investors who are familiar with anomalous January return for past winners and losers. In order to profit from the January effect, these investors may trade in the opposite direction — initiating buy trades for past losers in December and deferring buy trades for past winner to January.

We examine these competing hypotheses directly by using order imbalances. Specifically, we fit regression (3) separately in the months of December and January and also compare the slope coefficients on lagged returns for these months with those in the other months. Panel A of Table 3 reports the coefficients only for December and January. We find that all slope coefficients on lagged returns for OIB in the month of December are greater than the corresponding slope coefficients on lagged returns for both OIBNUM and OIBSH in all months in Table 3. For example, with OIBNUM as the dependent variable, the slope coefficients on one-month lagged return are −11.09 and −5.64, respectively, for all months and for December. The corresponding 12-month lagged return coefficients are 1.32 and 2.77. The one-month lagged return coefficients with OIBSH as dependent variable are −4.48 and −1.29 in all months and in December.

We also find that the most of the lagged-return coefficients in January are smaller than the corresponding ones in all months. For example, the one-month lagged return coefficient in all months is −11.09 compared with that of −15.88 in January when we use OIBNUM, and the corresponding coefficients are −4.48 and −8.54 with OIBSH.

To examine whether the coefficient estimates are statistically different in December and January compared to those in other months we fit the following regression:

$$b_{t,j} = b_{0,j} + b_{Dec,j} I_t \left(Dec\right) + b_{Jan,j} I_t \left(Jan\right) , \tag{5}$$

where $b_{t,j}$ is the estimated slope coefficient on lagged return at lag j in month t and *I(Dec)* and *I(Jan)* are the dummy variables for December and January, respectively. The slope coefficients $b_{Dec,j}$ and $b_{Jan,j}$ are the differences between the slope coefficients in December and January from the averages in other months. We use this regression to conduct our statistical tests allowing for the possibility that the slope coefficient estimates in December and January are correlated with the corresponding estimates in other months. To allow for the correlation, we compute the Newey-West standards errors with three lags.

Table 4 reports the coefficients of equation (4) and the corresponding *t*-statistics. The results indicate that for OIBNUM, the coefficient $b_{Dec,j}$ is positive for all lagged returns, and seven of 12 coefficients are statistically significant. With OIBSH, $b_{Dec,j}$ is positive for all lagged return coefficients, although only two of the coefficients are statistically significant. We can reject the hypothesis that $b_{Dec,j}$ for all lagged return coefficients are jointly equal to zero using the Bonferroni test. The positive coefficients indicate investors are more likely to sell losers and less likely to sell winners in December than in other months. However, the positive coefficients could also be due to investors higher propensity to buy winners as well. To ensure that the positive coefficients are consistent with a higher likelihood of selling losers, we estimate regression equation (3) with number of shares sold as a fraction of shares outstanding as the dependent variable (instead of OIB). The negative coefficient estimates for the December dummy from the corresponding regression (4) is indeed consistent with an increase in selling of losers in December as compared to other months.

The magnitude of the coefficients is economically significant. Following the same methodology as in Section III.C., we find that the average difference between extreme deciles, formed only for the month of December is 20.89% for OIBNUM and 17.60% for OIBSH. Therefore, our results provide support for the seasonal tax-loss selling hypothesis in December.

When we consider January, all lagged-return coefficients on January dummy are negative with OIBNUM, and half of them are statistically significant. With OIBSH, we find that five of the 12 coefficients are significantly negative. With the Bonferroni tests we can reject the hypothesis that $b_{Jan,j}$ for lagged returns are jointly equal to zero for both OIBSH and OIBNUM. Again, the coefficients are economically significant. The average difference between extreme deciles, formed only for the month of January, is 18.96% for OIBNUM and 16.07% for OIBSH. These results indicate that investors tend to sell more past winners in January than in other months to defer the realization of capital gains to the following calendar year.

The results for January are also consistent with more buying of losers as compared to other months. So, once again, we directly examine selling by estimating regression (3) with number of shares sold as a fraction of shares outstanding as the dependent variable. The coefficient estimate for the dummy variable for January from regression (4) is positive and is consistent with increased selling of winners in January as compared to the other months.

F. Results in perspective

Overall, our cross-sectional tests show a pattern of relation between investors' trading decisions and past returns that cannot be explained by any single hypothesis. Our evidence strongly supports the momentum trading hypothesis for four to 12-month horizons. We also find strong support for the seasonal tax-loss selling hypothesis.

The relation between short-horizon returns and order imbalances provide support for both the disposition effect and for contrarian trading. Since these two hypotheses make similar predictions about the sign of the slope coefficients, it is not possible to differentiate between them based solely on the signs. Generally, behavioral hypotheses such as the disposition effect do not offer sharp guidance on the length of the horizon over which the effect should be observed. However, if disposition effect were the dominant driving force, then we would expect the horizon over which we see the negative coefficients to be roughly in line with the length of the holding periods for investors potentially exposed to this effect.

The average holding period for individual investors is about one year and the holding period for mutual funds is also around one year.[12] The evidence that the negative coefficients are observed only at much shorter horizons indicates that the primary reason for the negative coefficient is unlikely to be the disposition effect for these investors. Moreover, the evidence that

[12] Wermers (2000) reports an average annual turnover for mutual funds of 70% during 1990–1994. Over the sample period 1980–2009, Huang, Sialm, and Zhang (2011) report an average turnover of 90%, which translates to a holding period of about 1.1 years.

the coefficients rapidly decline in magnitude from −11.09 for the one-month lagged return to −3.45 for two-month lagged return indicate that shorter term returns are more important in these investors' trading decisions. Therefore, it seems likely that active short-term traders who tend to be contrarians contribute more towards the shorter horizon negative coefficients than the disposition traders.

The results are also consistent with active traders exploiting the pattern of stock return predictability. The evidence in Jegadeesh (1990) indicates that at short horizons stocks exhibit return reversals, and hence investors may be seeking to exploit this empirical regularity. Jegadeesh also documents a return continuation at longer horizon and we find a positive relation between order imbalances and longer-lagged returns.

IV. Additional Tests

A. Evidence within subsamples

This subsection examines order imbalances separately for different subsamples. First we examine order imbalances separately for small and large firms. Each month, we classify a stock as a small stock if the market capitalization of equity at the end of the previous month is below sample median, and a big stock otherwise. These results are reported in Table 5 and are largely similar to the results for the full sample reported in Table 3. For instance, the OIBNUM regression coefficient on last-month return is −11.37 for large firms and −11.00 for small firms (cf. −11.09 for all firms). There is no significant difference between any coefficients between small and large firms. The adjusted-R^2 is higher for the large firm regressions perhaps because of higher persistence of their order imbalances.

Next, we divide the sample into high and low volatility periods based on realized aggregate market volatility in that month. We aggregate the Fama-Macbeth coefficients from equations (2) and (3) separately for these two sub-periods and analyze differences between them. In unreported results (available upon request), we do not find much difference in the coefficients on lagged returns after the first month. The OIBNUM coefficient on the first month lagged return is, however, less negative (albeit, statistically insignificantly so) in high volatility months than that in low volatility months. This suggests that contrarian trading by retail investors (buying stocks whose prices have recently declined) is less pronounced in more volatile time periods. We also sort the sample into high and low liquidity months based on aggregate Amihud (2002) liquidity measure and find that OIBNUM regression coefficients on the first two lags of returns are significantly more negative in periods of high illiquidity than those in periods of low illiquidity. This suggests that contrarian trading by retail investors is more pronounced when markets are illiquid.

B. Large orders

In our analysis, OIBNUM equally weights all trades and OIBSH weights order imbalances by the number of shares traded. Hence OIBSH is more representative of large investors while OIBNUM is more representative of small traders. To more directly examine the order imbalances of large traders, we compute OIBNUM and OIBSH using only trades that are larger than $10,000.[13]

Table 6 presents the results of regression (3) fitted with the large order subsample. The results are generally similar to the results in Table 3. However, the point estimates are smaller in magnitude in Table 6 than the corresponding estimates in Table 3. Consistent with our earlier observations, these results indicate that while past returns influence the trades of large traders, they do so more for small investors.

C. Pre-1997 Nasdaq data

As we discussed earlier, it is possible that some trader initiators use limit orders to execute their trades. These trade initiators may trade against other trade initiators but we will not be able to algorithmically identify that active traders were involved on both sides of such trades. While such crossing trades are not uncommon today, in the pre-1997 period Nasdaq market makers were typically on the passive side of each trade. Therefore, we can more accurately identify the trade initiator in the pre-1997 Nasdaq, and test the robustness of our findings.

We obtain Nasdaq trade and quote data from two sources. We obtain the Nasdaq data for the 1993 to 1996 period from TAQ, and for the 1987 to 1992 period from the Institute for the Study of Security Markets (ISSM). Figure 3 shows the value-weighted average of the order imbalances measured in shares and in number of trades. In this sample, the order imbalance measured in number of trades is lower than that measured in shares almost throughout the entire

[13] Chordia, Roll, and Subrahmanyam (2011) use the $10,000 cutoff to delineate large and small trades.

sample period, which is similar to the pattern we observed for NYSE stocks in the early part of the sample period.

Table 7 presents the results of regression (3) fitted with pre-1997 Nasdaq data. The results in Table 7 are quite similar to the results in Table 3. For example, the slope coefficients on first three lagged returns are significantly negative in both tables, and the coefficients at longer lags are all positive, and most of them are statistically significant. Therefore, these results also indicate that investors are contrarian traders based on short-horizon returns but momentum traders based on longer horizon returns.

The results for December and January also are similar to the corresponding results in Table 3. The slope coefficients in December are generally more positive, and the slope coefficients in January are generally less positive than the corresponding coefficients in other months. Therefore, the Nasdaq results also support the seasonal tax-motivated trading hypothesis. These results indicate that our findings with data from the more recent period are not driven by possible misclassification of the side that initiated the trade.

V. Conclusions

We examine the relation between order imbalances and past returns and test a number of hypotheses about motivations for trading that are proposed in the literature. Theoretical models of trading typically identify whether it is buyers or sellers who initiate trades under various circumstances. Therefore, empirical tests of these models must take into account the initiators of the trade. Empirical tests using turnover data do not identify the initiators and they could lead to misleading inferences. In many instances, it is hard to differentiate among various hypotheses using turnover data, while we are able to do so using order imbalances. Our paper more directly tests the theoretical models by explicitly considering whether buyers or sellers initiate trades.

In our time-series tests, we find a negative relation between aggregate market order imbalances and one-month lagged returns. We do not find a significant relation between order imbalances and longer horizon returns. Our time-series tests do not find any relation between various measures of market risk and order imbalances. Therefore, our results do not provide support for the flight-to-safety hypothesis.

We find a far stronger relation between order imbalances and past returns in our cross-sectional tests. These tests show a pattern of relation between investors' trading decisions and past returns that cannot be fully explained by any single hypothesis. Our evidence strongly supports the momentum trading hypothesis for four to 12-month horizons. We also find strong support for the seasonal tax-loss selling hypothesis.

We find a negative relation between short-horizon returns and order imbalances, which is consistent with both the disposition effect and the contrarian trading hypothesis. If disposition effect were the dominant driving force, then we would expect the horizon over which we see the negative coefficients to be roughly in line with the length of the holding periods for investors potentially exposed to this effect. However, we observe the negative relation over a horizon of up to three months, which is shorter than the typical holding period of individuals and institutions.

In addition to past returns, we find that market capitalization, book to market ratio and past order imbalances are useful in predicting future order imbalances. Our empirical model can help in identifying the expected and unexpected components of order flow, and future work can examine how these components are related to market impact and price discovery. Also, investors can potentially use the predictable pattern of order flow that we document to develop efficient trade execution strategies.

Appendix

This appendix reports the results of regression (A1) below, that excludes lagged order imbalances from the right-hand side of the regression (3). Specifically, we fit the regression:

$$OIB_{i,t} = a_t + \sum_{j=1}^{12} b_{t,j} R_{i,t-j} + c_t \ln\left(Size_{i,t-1}\right) + d_t \ln\left(BM_{i,t-1}\right) + e_t Vol_{i,t-1} + u_{i,t}. \tag{A1}$$

Table A1 reports the results of this regression. When OIBNUM is the dependent variable, the average adjusted-R^2 reduces from 32.0% for regression (3) to 7.5% for regression (A1) and it reduces similarly for OIBSH as well. Therefore, a large part of the explanatory power in the regression comes from the inclusion of order imbalances.

For the regression with OIBNUM, the slope coefficients on lagged returns are significantly negative at the shorter lags and significantly positive for longer lags; the results are, thus, similar for regressions (3) and (A1). For the regression with OIBSH, the slope coefficients at longer returns are significantly positive, but the only the coefficient on two-month lagged return is significantly negative.

The likely reason for the difference between the results of regressions (3) and (A1) is the omitted variable bias. For instance, one-month lagged return is uncorrelated with all other explanatory variables in regression (3) except with OIB_{t-1}. When OIB_{t-1} is excluded from the regression, the slope coefficient on one-month lagged returns partly absorbs the positive relation between OIB_t and OIB_{t-1} and hence the slope coefficient in regression (A1) is less negative than that in regression (3). Therefore, it is important to include lagged order imbalances in the regressions.

References

Admati, Anat R., and Paul Pfleiderer, 1988, "A Theory of Intraday Patterns: Volume and Price Variability," *Review of Financial Studies* 1, 3–40.

Amihud, Yakov, 2002, "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects," *Journal of Financial Markets* 5, 31–56.

Beber, Alessandro, Michael Brandt, and Kenneth Kavajecz, 2009, "Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market," *Review of Financial Studies* 22, 925–957.

Ben-David, Itzhak, and David Hirshleifer, 2012, "Are Investors Really Reluctant to Realize Their Losses? Trading Responses to Past Returns and the Disposition Effect," *Review of Financial Studies* 25, 2485–2532.

Brennan, Michael J., and Henry Cao, 1997, "International Portfolio Investment Flows," *Journal of Finance* 52, 1851–1880.

Caballero, Ricardo, and Arvind Krishnamurthy, 2008, "Collective Risk Management in a Flight to Quality Episode," *Journal of Finance* 53, 2195–2230.

Chakrabarty, Bidisha, Pamela C. Moulton, and Andriy Shkilko, 2012, "Short Sales, Long Sales, and the Lee-Ready Trade Classification Algorithm Revisited," *Journal of Financial Markets* 15, 467–491.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2001, "Market Liquidity and Trading Activity," *Journal of Finance* 56, 501–530.

Chordia Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2005, "Evidence on the Speed of Convergence to Market Efficiency," *Journal of Financial Economics* 76, 271–292.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2011, "Recent Trends in Trading Activity and Market Quality," *Journal of Financial Economics* 101, 243–263.

Connolly, Robert A., Christopher T. Stivers, and Licheng Sun, 2005, "Stock Market Uncertainty and the Stock-Bond Return Relation," *Journal of Financial and Quantitative Analysis* 40, 161–194.

Constantinides, George, 1984, "Optimal Stock Trading with Personal Taxes: Implications for Prices and the Abnormal January Returns," *Journal of Financial Economics* 13, 65–89.

Coval, Joshua D., David Hirshleifer, and Tyler Shumway, 2005, "Can Individual Investors Beat the Market?" Working paper.

Ellis, Katrina, 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.

Fama, Eugene F., and Kenneth R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance* 47, 427–465.

Fama, Eugene F., and James D. MacBeth, 1973, "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy* 81, 607–636.

Griffin, John M., Fredrico Nardari, and René M. Stulz, 2007, "Do Investors Trade More When Stocks Have Performed Well? Evidence from 46 Countries," *Review of Financial Studies* 20, 905–952.

Hong, Harrison, and Jeremy C. Stein, 1999, "A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets," *Journal of Finance* 54, 2143–2184.

Huang, Jennifer, Clemens Sialm, and Hanjiang Zhang, 2011, "Risk Shifting and Mutual Fund Performance," *Review of Financial Studies* 24, 2575–2616.

Jegadeesh, Narasimhan, 1990, "Evidence of Predictable Behavior of Security Returns," *Journal of Finance* 45, 881–898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, "Returns to Buying Winners and Selling Losers: Implications for Market Efficiency," *Journal of Finance* 48, 65–91.

Kahneman, Daniel, and Amos Tversky, 1979, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica* 47, 263–291.

Kyle, Albert S., 1985, "Continuous Auctions and Insider Trading," *Econometrica* 53, 1315–1335.

Lakonishok, Josef, and Seymour Smidt, 1986, "Volume for Winners and Losers: Taxation and Other Motives for Trading," *Journal of Finance* 41, 951–974.

Lee, Charles M.C., and Balkrishna Radhakrishna, 2000, "Inferring Investor Behavior: Evidence from TORQ Data," *Journal of Financial Markets* 3, 88–111.

Lee, Charles M.C., and Mark J. Ready, 1991, "Inferring Trade Direction from Intraday Data," *Journal of Finance* 46, 733–746.

Madhavan, Ananth, Kewei Ming, Vesna Strasser, and Yingchuan Wang, 2002, "How Effective are Effective Spreads? An Evaluation of Trade Side Classification Algorithms," Working paper, ITG Inc.

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* 52, 1775–1798.

Orosel, Gerhard O., 1998, “Participation Costs, Trend Chasing, and Volatility of Stock Prices,” *Review of Financial Studies* 11, 521–557.

Pástor, Luboš, and Robert F. Stambaugh, 2003, “Liquidity Risk and Expected Stock Returns,” *Journal of Political Economy* 111, 642–685.

Poterba, James M., and Scott J. Weisbenner, 2001, “Capital Gains Tax Rules, Tax-loss Trading, and Turn-of-the-year Returns,” *Journal of Finance* 56, 353–368.

Sarkar, Asani, and Robert A. Schwartz, 2009, “Market Sidedness: Insights into Motives for Trade Initiation,” *Journal of Finance* 54, 375–423.

Shefrin, Hersh, and Meir Statman, 1985, “The Disposition to Sell Winners too Early and Ride Losers too Long: Theory and Evidence,” *Journal of Finance* 40, 777–791.

Statman, Meir, Steven Thorley, and Keith Vorkink, 2006, “Investor Overconfidence and Trading Volume,” *Review of Financial Studies* 19, 1531–1565.

Wang, Jiang, 1993, “A Model of Intertemporal Asset Prices Under Asymmetric Information,” *Review of Economic Studies* 60, 249–282.

Wermers, Russ, 2000, “Mutual Fund Performance: An Empirical Decomposition into Stock-picking Talent, Style, Transactions Costs, and Expenses,” *Journal of Finance* 55, 1655–1695.

Figure 1: Order imbalance over time for NYSE stocks

This figure shows the value-weighted average (using market capitalization as weights) of order imbalance over time. OIBNUM is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. The sample includes all NYSE stocks over the period 1993 to 2010.

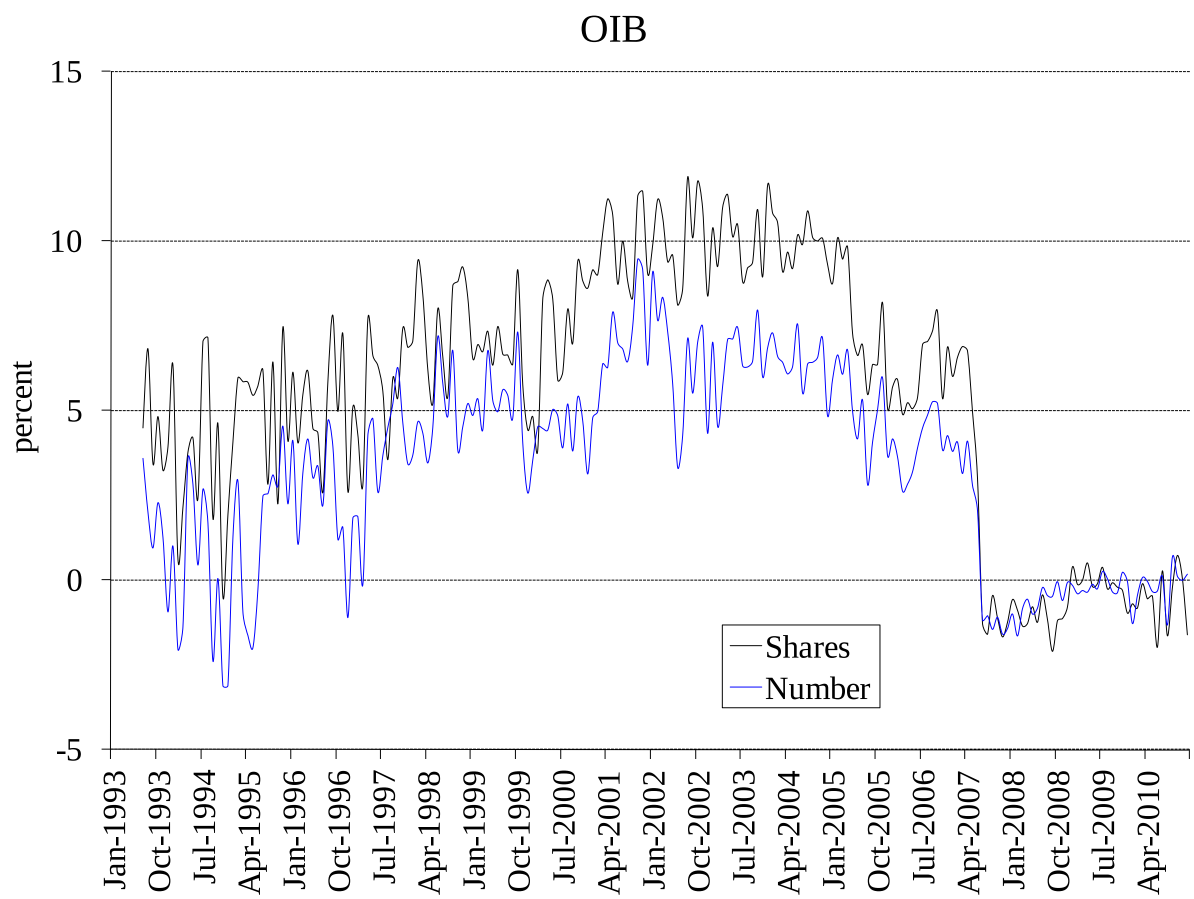
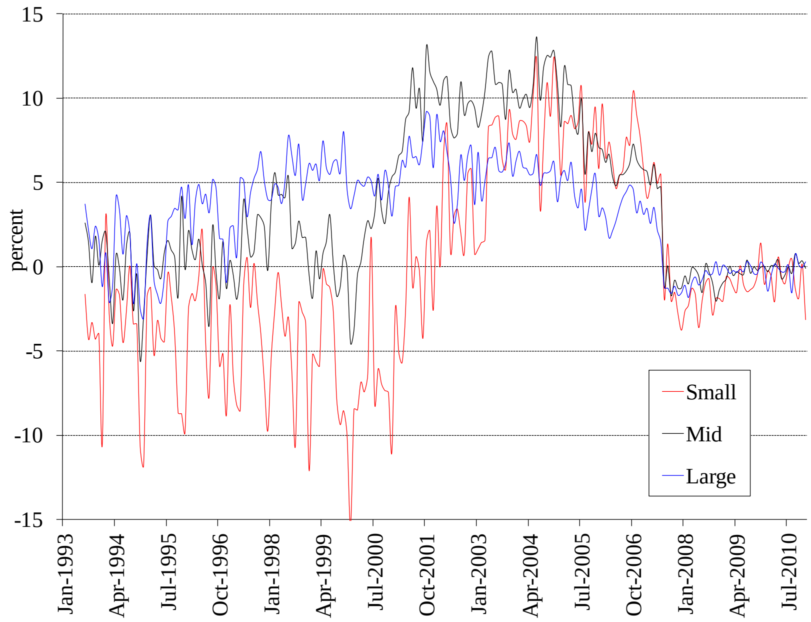


Figure 2: Order imbalance over time for different market capitalization NYSE stocks

This figure shows the value-weighted average (using market capitalization as weights) of order imbalance over time. OIBNUM (Panel A) is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH (Panel B) is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. Stocks are divided into quintiles based on their market capitalization (using breakpoints of the stocks in the sample). The figure reports the order imbalance for only three size categories—small, medium, and large. The sample includes all NYSE stocks over the period 1993 to 2010.

Panel A: OIBNUM



Panel B: OIBSH

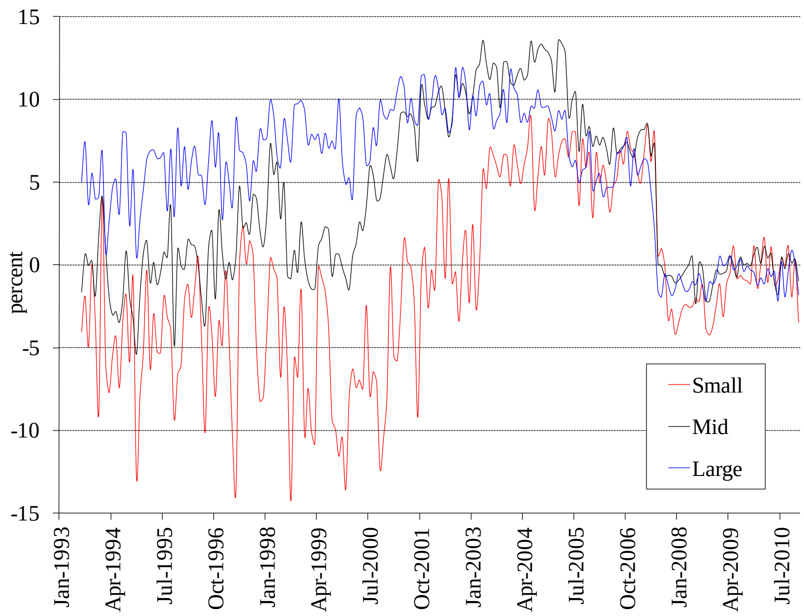


Figure 3: Order imbalance over time for Nasdaq stocks

This figure shows the value-weighted average (using market capitalization as weights) of order imbalance over time. OIBNUM is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. The sample includes all Nasdaq stocks over the period 1987 to 1996.

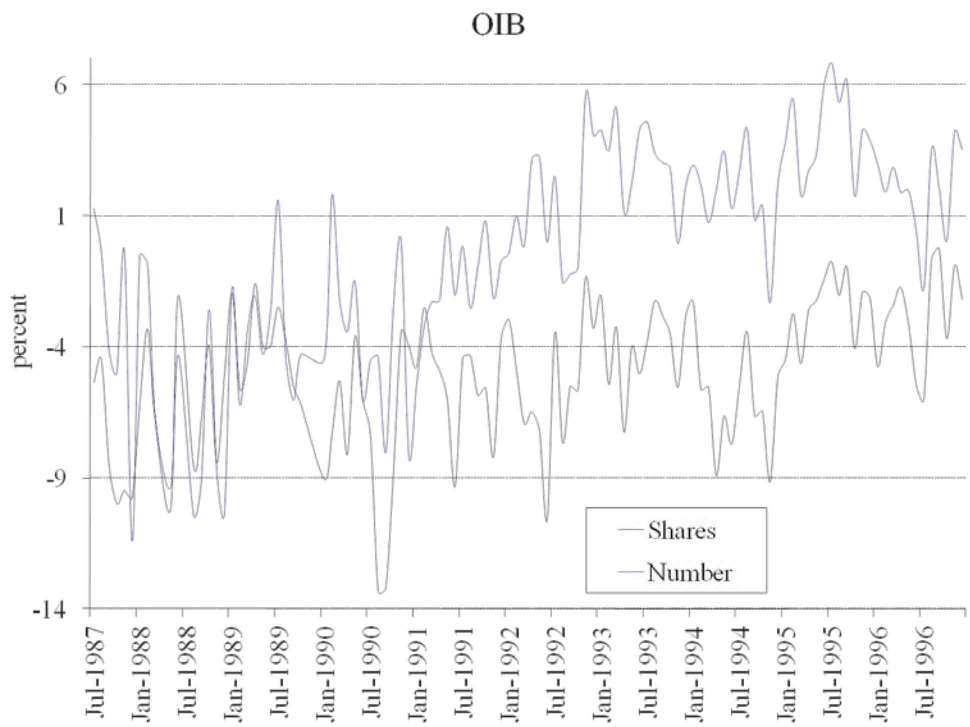


Table 1: Descriptive statistics for NYSE stocks

This table presents descriptive statistics on the key variables. OIBNUM is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. Both measures of OIB are reported in percent per month. Return is in percent per month, Market cap is market capitalization in billions of dollars, turnover is in percent per month, and volatility is the standard deviation of returns in percent per month calculated using daily returns within the month. All summary statistics are reported as the time-series averages of monthly cross-sectional measures. Each month, a stock is classified as a small (big) stock if its market capitalization is lower (higher) than the median market capitalization that month. Each month, a stock is classified as a growth (value) stock if its book-to-market is lower (higher) than the median book-to-market that month. Panels B through E report the statistics for these subsamples of firms. The sample includes all NYSE stocks over the period 1993 to 2010.

	Mean	Median	StdDev	Skewness	Kurtosis
Panel A: All stocks					
OIBNUM	2.474	3.315	10.526	-0.718	7.849
OIBSH	2.990	4.225	13.654	-0.729	7.889
Turnover	13.767	10.130	14.303	5.406	75.461
Return	1.240	0.656	12.052	1.194	18.112
Market cap	5.367	1.106	17.293	8.756	109.157
Book-to-market	0.797	0.543	1.026	4.148	27.529
Volatility	12.062	10.272	8.218	4.868	71.118
Panel B: Small stocks					
OIBNUM	1.191	2.026	12.212	-0.582	5.666
OIBSH	0.248	1.341	16.211	-0.503	5.717
Turnover	12.689	8.695	15.196	5.176	60.526
Return	0.918	0.112	14.109	1.163	15.207
Market cap	0.432	0.378	0.306	0.500	2.137
Book-to-market	1.088	0.712	1.322	3.336	17.295
Volatility	14.184	12.125	9.534	3.920	42.148
Panel C: Big stocks					
OIBNUM	3.648	4.033	8.167	-0.270	6.800
OIBSH	5.406	5.692	9.861	-0.220	6.611
Turnover	14.833	11.332	12.959	4.400	43.049
Return	1.559	1.083	9.316	0.812	10.024
Market cap	10.300	3.350	23.432	6.423	59.195
Book-to-market	0.518	0.421	0.444	3.664	35.951
Volatility	9.903	8.894	4.805	2.312	16.909
Panel D: Growth stocks					
OIBNUM	4.170	4.598	8.738	-0.442	7.727
OIBSH	4.776	5.447	11.049	-0.523	7.257
Turnover	14.965	11.290	13.742	4.283	42.500
Return	2.200	1.482	10.765	1.118	12.494
Market cap	8.375	1.954	23.003	6.858	66.100
Book-to-market	0.315	0.322	0.135	-0.220	2.089
Volatility	11.110	9.880	5.744	2.618	20.327
Panel E: Value stocks					
OIBNUM	1.008	1.828	11.314	-0.631	6.335
OIBSH	1.536	2.799	14.994	-0.617	6.610
Turnover	12.383	9.053	13.170	4.734	53.983
Return	0.243	-0.171	12.252	0.850	14.413
Market cap	2.873	0.635	8.412	8.065	99.805
Book-to-market	1.278	0.882	1.260	3.673	19.361
Volatility	12.413	10.494	8.412	3.757	37.248

Table 2: Time-series determinants of order imbalance for NYSE stocks

We form value-weighted size sorted quintile portfolios and regress order imbalances (OIB) on past twelve months of returns, the past month's portfolio volatility, and three lags of OIB as follows:

$$OIB_{i,t} = a_i + \sum_{j=1}^{12} b_{i,j} R_{i,t-j} + c_i Vol_{i,t-1} + \sum_{j=1}^{3} d_{i,j} OIB_{i,t-j} + e_{i,t} \; .$$

OIBNUM is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. Return is in percent per month and Vol is the standard deviation of returns in percent per month calculated using daily returns within the month. Numbers within parentheses next to the coefficient estimate are the Newey-West corrected *t*-statistics using three lags. The sample includes all NYSE stocks over the period 1993 to 2010.

	1=Small	2	3	4	5=Big	All
			Panel A: Order imbalance in number of trades, OIBNUM			
CNST	0.00 (-0.01)	0.00 (0.74)	0.00 (0.59)	0.00 (0.08)	0.00 (0.39)	0.00 (0.50)
Return(-1)	-0.01 (-0.55)	-0.08 (-3.08)	-0.07 (-3.26)	-0.06 (-2.93)	-0.07 (-2.66)	-0.07 (-3.00)
Return(-2)	-0.03 (-0.90)	0.01 (0.22)	0.00 (0.01)	0.01 (0.41)	0.01 (0.36)	0.00 (0.19)
Return(-3)	-0.05 (-1.54)	-0.08 (-3.26)	-0.06 (-2.58)	-0.01 (-0.58)	0.02 (1.07)	0.01 (0.55)
Return(-4)	0.01 (0.21)	0.05 (1.89)	0.02 (0.81)	0.03 (1.19)	-0.01 (-0.58)	-0.01 (-0.45)
Return(-5)	0.01 (0.49)	0.02 (0.90)	0.02 (1.14)	0.03 (1.52)	0.05 (2.17)	0.05 (2.27)
Return(-6)	0.00 (-0.00)	0.03 (0.95)	0.02 (0.81)	0.02 (0.83)	0.01 (0.53)	0.00 (0.17)
Return(-7)	0.01 (0.36)	-0.04 (-1.63)	-0.02 (-1.08)	-0.03 (-1.37)	0.00 (-0.17)	0.00 (-0.24)
Return(-8)	-0.02 (-0.53)	0.01 (0.52)	0.00 (-0.11)	-0.02 (-0.99)	-0.01 (-0.62)	-0.02 (-0.83)
Return(-9)	-0.01 (-0.26)	-0.03 (-1.05)	-0.02 (-0.81)	-0.01 (-0.29)	0.00 (-0.16)	-0.01 (-0.21)
Return(-10)	0.01 (0.37)	0.00 (0.13)	-0.01 (-0.19)	0.00 (0.19)	0.01 (0.30)	0.00 (0.11)
Return(-11)	-0.02 (-0.53)	-0.01 (-0.28)	-0.01 (-0.53)	-0.01 (-0.27)	0.00 (-0.09)	0.00 (-0.17)
Return(-12)	0.02 (0.55)	0.01 (0.40)	0.03 (1.14)	0.02 (1.28)	0.02 (1.21)	0.02 (1.26)
Volatility(-1)	0.02 (0.45)	0.00 (0.04)	0.01 (0.12)	0.02 (0.49)	0.02 (0.34)	0.00 (0.08)
OIB(-1)	0.56 (7.93)	0.66 (8.82)	0.61 (6.97)	0.63 (6.35)	0.61 (7.06)	0.62 (7.30)
OIB(-2)	0.12 (1.49)	0.11 (0.99)	0.20 (1.91)	0.09 (0.75)	0.13 (1.21)	0.12 (1.08)
OIB(-3)	0.24 (3.50)	0.17 (2.09)	0.16 (1.88)	0.25 (2.71)	0.19 (2.48)	0.20 (2.50)
$\bar{R}^2$ (%)	71.6	78.6	87.0	88.2	77.8	80.3

	1=Small	2	3	4	5=Big	All
			Panel B: Order imbalance in shares traded, OIBSH			
CNST	-0.01 (-2.68)	-0.01 (-2.28)	0.00 (0.58)	0.03 (5.47)	0.03 (3.34)	0.03 (3.23)
Return(-1)	0.07 (2.24)	-0.01 (-0.32)	-0.03 (-1.26)	-0.06 (-1.81)	-0.09 (-2.13)	-0.09 (-2.35)
Return(-2)	0.00 (-0.14)	0.02 (0.49)	0.00 (-0.13)	-0.03 (-0.99)	-0.06 (-1.49)	-0.07 (-1.81)
Return(-3)	-0.05 (-1.45)	-0.04 (-1.87)	-0.04 (-1.50)	-0.03 (-1.19)	-0.06 (-1.67)	-0.06 (-1.87)
Return(-4)	-0.01 (-0.16)	0.07 (3.29)	0.04 (1.77)	-0.01 (-0.34)	-0.04 (-0.91)	-0.03 (-0.98)
Return(-5)	0.03 (0.98)	0.03 (1.33)	0.02 (0.82)	0.00 (-0.01)	0.01 (0.30)	0.01 (0.30)
Return(-6)	-0.03 (-0.83)	0.01 (0.51)	0.00 (0.04)	-0.01 (-0.47)	-0.01 (-0.35)	-0.01 (-0.50)
Return(-7)	0.09 (2.44)	0.00 (-0.06)	0.01 (0.32)	-0.01 (-0.57)	0.00 (0.12)	0.00 (-0.05)
Return(-8)	-0.04 (-0.86)	-0.01 (-0.29)	-0.01 (-0.49)	-0.01 (-0.52)	-0.02 (-0.63)	-0.02 (-0.82)
Return(-9)	0.00 (-0.01)	0.00 (-0.14)	0.00 (0.11)	0.00 (0.04)	-0.03 (-0.70)	-0.02 (-0.59)
Return(-10)	0.04 (1.41)	0.02 (0.84)	-0.01 (-0.25)	-0.02 (-0.80)	-0.01 (-0.17)	-0.01 (-0.30)
Return(-11)	-0.04 (-1.10)	-0.02 (-0.71)	-0.02 (-1.07)	-0.01 (-0.44)	-0.02 (-0.57)	-0.02 (-0.60)
Return(-12)	0.02 (0.46)	0.02 (0.73)	0.02 (0.92)	0.00 (-0.09)	-0.01 (-0.33)	-0.01 (-0.30)
Volatility(-1)	0.09 (1.65)	0.11 (1.99)	0.02 (0.31)	-0.19 (-2.78)	-0.24 (-2.61)	-0.25 (-2.49)
OIB(-1)	0.44 (5.85)	0.62 (8.58)	0.64 (7.82)	0.73 (5.65)	0.58 (4.26)	0.60 (4.61)
OIB(-2)	0.14 (1.66)	0.15 (1.90)	0.15 (1.65)	0.14 (1.16)	0.27 (2.32)	0.27 (2.27)
OIB(-3)	0.23 (3.56)	0.19 (2.69)	0.23 (2.67)	0.17 (1.61)	0.36 (3.15)	0.31 (2.89)
$\bar{R}^2$ (%)	66.8	79.3	85.5	82.8	69.0	73.4

Table 3: Cross-sectional determinants of order imbalance for NYSE stocks

We run the following cross-sectional regression each month:

$$OIB_{i,t} = a_t + \sum_{j=1}^{12} b_{t,j} R_{i,t-j} + c_t \ln\left(Size_{i,t-1}\right) + d_t \ln\left(BM_{i,t-1}\right) + e_t Vol_{i,t-1} + \sum_{j=1}^{3} f_{t,j} OIB_{i,t-j} + u_{i,t}.$$

The table reports the time-series averages of the coefficients (multiplied by 100) together with their Newey-West corrected *t*-statistics (using three lags) within parentheses. OIBNUM is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. Return is in percent per month, Size is market capitalization in millions of dollars, BM is book-to-market, and Vol is the standard deviation of returns in percent per month calculated using daily returns within the month. The full sample in Panel A includes all NYSE stocks over the period 1993 to 2010. Panels B and C present the results for all months for sample periods of 1993 to 2001 and 2002 to 2010, respectively.

	OIBNUM	OIBSH	OIBNUM	OIBSH	OIBNUM	OIBSH
	Panel A: Full sample period (1993–2010)					
	All months		December		January	
Constant	-1.79 (-2.35)	-11.55 (-8.06)	-4.72 (-1.40)	-12.88 (-2.67)	-0.41 (-0.28)	-9.31 (-3.47)
Return(-1)	-11.09 (-7.38)	-4.48 (-10.08)	-5.64 (-2.01)	-1.29 (-1.71)	-15.88 (-2.29)	-8.54 (-4.59)
Return(-2)	-3.45 (-7.58)	-3.77 (-10.30)	0.26 (0.15)	-2.71 (-1.53)	-8.17 (-2.16)	-4.57 (-3.54)
Return(-3)	-1.33 (-4.52)	-2.48 (-6.38)	1.86 (1.07)	-0.57 (-0.25)	-5.75 (-3.39)	-5.77 (-6.65)
Return(-4)	1.10 (3.68)	0.43 (1.24)	3.01 (1.58)	1.56 (0.90)	-1.97 (-1.08)	-3.37 (-2.42)
Return(-5)	1.79 (5.62)	1.17 (4.01)	3.46 (2.08)	2.21 (2.10)	-1.90 (-2.10)	-0.43 (-0.44)
Return(-6)	1.42 (4.61)	1.45 (4.12)	5.36 (2.37)	3.31 (3.01)	0.12 (0.25)	-0.95 (-1.29)
Return(-7)	1.15 (4.78)	0.92 (3.12)	4.32 (2.67)	2.99 (2.26)	-2.36 (-2.61)	-2.50 (-3.65)
Return(-8)	1.17 (4.53)	0.84 (2.80)	5.15 (2.74)	2.35 (2.13)	-1.94 (-1.88)	0.37 (0.55)
Return(-9)	0.62 (1.78)	1.03 (3.13)	2.46 (1.22)	2.49 (2.63)	-0.60 (-1.04)	-0.12 (-0.17)
Return(-10)	0.75 (2.84)	0.49 (1.92)	4.25 (2.12)	2.58 (2.57)	-0.88 (-1.84)	-0.71 (-0.73)
Return(-11)	1.08 (4.20)	0.34 (1.09)	1.55 (1.20)	1.67 (1.02)	0.20 (0.50)	2.18 (1.53)
Return(-12)	1.32 (5.07)	1.39 (4.57)	2.77 (1.81)	2.60 (2.06)	0.28 (0.36)	1.50 (1.58)
ln(Size(-1))	0.17 (4.08)	0.90 (9.47)	0.32 (1.79)	0.95 (2.91)	0.10 (0.99)	0.79 (4.64)
ln(BM(-1))	-0.46 (-8.19)	-0.11 (-2.65)	-0.63 (-3.35)	-0.36 (-1.95)	-0.42 (-2.15)	0.07 (0.41)
Volatility(-1)	-3.55 (-4.17)	0.31 (0.39)	-16.42 (-2.33)	-7.18 (-2.90)	3.07 (1.08)	2.66 (0.89)
OIB(-1)	35.57 (22.76)	19.02 (18.15)	34.59 (6.05)	18.86 (6.64)	32.73 (9.69)	17.29 (6.84)
OIB(-2)	12.17 (14.94)	10.86 (13.16)	15.09 (10.10)	12.06 (6.65)	11.44 (5.42)	11.26 (5.54)
OIB(-3)	11.27 (16.82)	10.43 (16.09)	12.72 (5.62)	10.62 (6.42)	12.63 (9.95)	12.93 (9.99)
Average $\bar{R}^2$ (%)	32.0	15.9	36.5	18.3	31.3	15.7
Average number of stocks	1,388	1,388	1,374	1,374	1,381	1,381

	OIBNUM	OIBSH	OIBNUM	OIBSH
	Panel B: Sample 1993–2001		Panel C: Sample 2002–2010	
Constant	-4.60 (-4.79)	-20.78 (-18.30)	0.68 (0.75)	-3.43 (-3.39)
Return(-1)	-20.67 (-12.12)	-4.85 (-5.89)	-2.66 (-9.55)	-4.15 (-10.20)
Return(-2)	-5.47 (-7.70)	-3.66 (-6.38)	-1.67 (-5.58)	-3.88 (-8.25)
Return(-3)	-1.32 (-2.40)	-1.97 (-3.73)	-1.33 (-4.93)	-2.93 (-5.31)
Return(-4)	2.29 (4.85)	1.02 (1.71)	0.05 (0.21)	-0.09 (-0.25)
Return(-5)	3.15 (6.55)	1.77 (3.56)	0.58 (2.32)	0.64 (2.20)
Return(-6)	2.57 (4.95)	2.39 (4.11)	0.41 (1.78)	0.62 (1.74)
Return(-7)	1.97 (5.16)	1.48 (3.02)	0.43 (1.87)	0.43 (1.33)
Return(-8)	1.98 (4.61)	1.17 (2.32)	0.46 (1.91)	0.55 (1.59)
Return(-9)	1.60 (2.67)	1.73 (3.39)	-0.25 (-0.80)	0.41 (1.05)
Return(-10)	1.65 (3.66)	0.95 (2.34)	-0.04 (-0.20)	0.09 (0.29)
Return(-11)	1.88 (4.83)	0.67 (1.36)	0.38 (1.41)	0.05 (0.13)
Return(-12)	2.61 (7.61)	2.23 (4.20)	0.19 (0.86)	0.65 (2.58)
ln(Size(-1))	0.32 (5.20)	1.51 (23.21)	0.05 (0.99)	0.36 (4.85)
ln(BM(-1))	-0.77 (-10.16)	-0.05 (-0.66)	-0.20 (-5.36)	-0.16 (-4.25)
Volatility(-1)	-2.90 (-1.84)	3.05 (2.35)	-4.11 (-5.17)	-2.09 (-2.85)
OIB(-1)	43.23 (42.35)	17.17 (25.54)	28.83 (13.79)	20.65 (11.31)
OIB(-2)	13.34 (21.39)	9.37 (14.80)	11.14 (7.94)	12.18 (8.68)
OIB(-3)	11.86 (23.00)	9.94 (18.40)	10.75 (9.20)	10.87 (9.72)
Average $\bar{R}^2$ (%)	40.5	13.7	24.6	17.8
Average number of stocks	1,500	1,500	1,289	1,289

Table 4: Tax-loss selling in NYSE stocks

We run the following cross-sectional regression each month:

$$OIB_{i,t} = a_t + \sum_{j=1}^{12} b_{t,j} R_{i,t-j} + c_t \ln\left(Size_{i,t-1}\right) + d_t \ln\left(BM_{i,t-1}\right) + e_t Vol_{i,t-1} + \sum_{j=1}^{3} f_{t,j} OIB_{i,t-j} + u_{i,t}.$$

We then run a second-stage regression on the time-series of $b_{t,j}$ coefficients as follows:

$$b_{t,j} = b_{0,j} + b_{Dec,j} I_t\left(Dec\right) + b_{Jan,j} I_t\left(Jan\right),$$

where the *I(.)* indicator variables are one if the corresponding month is December or January, and zero otherwise. OIBNUM is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. Return is in percent per month, Size is market capitalization in millions of dollars, BM is book-to-market, and Vol is the standard deviation of returns in percent per month calculated using daily returns within the month. The table reports the dummy coefficients b_{Dec} and b_{Jan} together with their Newey-West corrected *t*-statistics (using three lags) within parentheses. The sample includes all NYSE stocks over the period 1993 to 2010.

	OIBNUM	OIBSH
Return(-1)*Dec	5.54 (2.83)	3.12 (2.96)
Return(-2)*Dec	3.64 (2.90)	1.10 (0.80)
Return(-3)*Dec	3.08 (2.75)	1.79 (1.02)
Return(-4)*Dec	1.81 (1.57)	0.89 (0.59)
Return(-5)*Dec	1.50 (1.37)	1.00 (1.07)
Return(-6)*Dec	4.21 (3.03)	1.82 (1.64)
Return(-7)*Dec	3.15 (2.71)	1.95 (1.61)
Return(-8)*Dec	4.09 (3.52)	1.61 (1.56)
Return(-9)*Dec	1.92 (1.61)	1.50 (1.91)
Return(-10)*Dec	3.70 (3.06)	2.18 (2.19)
Return(-11)*Dec	0.43 (0.53)	1.64 (1.52)
Return(-12)*Dec	1.50 (1.55)	1.34 (1.13)
Return(-1)*Jan	-4.70 (-1.24)	-4.13 (-2.89)
Return(-2)*Jan	-4.80 (-2.22)	-0.76 (-0.59)
Return(-3)*Jan	-4.53 (-3.53)	-3.41 (-3.57)
Return(-4)*Jan	-3.17 (-1.91)	-4.04 (-2.74)
Return(-5)*Jan	-3.87 (-4.21)	-1.65 (-1.60)
Return(-6)*Jan	-1.02 (-1.33)	-2.44 (-2.75)
Return(-7)*Jan	-3.53 (-3.84)	-3.54 (-4.64)
Return(-8)*Jan	-3.00 (-3.22)	-0.37 (-0.37)
Return(-9)*Jan	-1.14 (-1.48)	-1.11 (-1.44)
Return(-10)*Jan	-1.43 (-1.95)	-1.11 (-1.18)
Return(-11)*Jan	-0.92 (-1.35)	2.14 (1.93)
Return(-12)*Jan	-0.99 (-1.40)	0.23 (0.26)

Table 5: Order imbalance for small and large capitalization NYSE stocks

We run the following cross-sectional regression each month separately for small and large capitalization stocks:

$$OIB_{i,t} = a_t + \sum_{j=1}^{12} b_{t,j} R_{i,t-j} + c_t \ln\left(Size_{i,t-1}\right) + d_t \ln\left(BM_{i,t-1}\right) + e_t Vol_{i,t-1} + \sum_{j=1}^{3} f_{t,j} OIB_{i,t-j} + u_{i,t}.$$

The table reports the time-series averages of the coefficients together with their Newey-West corrected *t*-statistics within parentheses. OIBNUM is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. Return is in percent per month, Size is market capitalization in millions of dollars, BM is book-to-market, and Vol is the standard deviation of returns in percent per month calculated using daily returns within the month. The sample includes all NYSE stocks over the period 1993 to 2010.

	OIBNUM	OIBSH	OIBNUM	OIBSH
	Big stocks		Small stocks	
Constant	2.52 (2.90)	-1.78 (-1.39)	-10.85 (-7.81)	-24.05 (-13.53)
Return(-1)	-11.37 (-7.76)	-6.13 (-9.72)	-11.00 (-7.50)	-4.21 (-7.79)
Return(-2)	-2.14 (-5.33)	-4.73 (-9.57)	-4.11 (-7.78)	-3.35 (-7.42)
Return(-3)	-0.34 (-0.84)	-3.85 (-8.49)	-2.07 (-6.51)	-1.86 (-3.70)
Return(-4)	1.59 (4.05)	0.32 (0.72)	0.36 (1.12)	-0.08 (-0.18)
Return(-5)	2.03 (5.16)	0.45 (1.39)	1.47 (4.36)	1.37 (3.06)
Return(-6)	1.35 (3.49)	0.46 (1.26)	1.09 (3.11)	1.54 (3.23)
Return(-7)	1.33 (3.66)	0.27 (0.70)	0.94 (3.65)	1.01 (2.48)
Return(-8)	1.19 (3.61)	0.43 (1.08)	0.84 (3.00)	0.66 (1.44)
Return(-9)	0.69 (2.11)	1.17 (3.71)	0.30 (0.71)	0.46 (1.02)
Return(-10)	1.05 (3.32)	0.96 (2.37)	0.34 (1.08)	-0.15 (-0.42)
Return(-11)	1.55 (4.42)	0.58 (1.53)	0.69 (1.98)	0.01 (0.01)
Return(-12)	1.31 (4.56)	1.27 (3.57)	0.94 (3.19)	0.90 (2.23)
ln(Size(-1))	-0.11 (-2.55)	0.26 (3.19)	0.90 (8.05)	1.85 (13.52)
ln(BM(-1))	-0.34 (-5.99)	0.08 (1.57)	-0.29 (-5.13)	-0.09 (-1.42)
Vol(-1)	-3.72 (-3.84)	0.90 (0.83)	-2.18 (-2.40)	3.25 (3.30)
OIB(-1)	39.55 (17.54)	22.12 (16.35)	32.29 (25.33)	16.52 (16.99)
OIB(-2)	11.29 (12.99)	11.85 (12.39)	11.19 (13.22)	9.19 (11.30)
OIB(-3)	12.09 (15.06)	12.39 (15.60)	9.90 (15.13)	8.33 (13.01)
Average $\overline{R}^2$ (%)	40.0	17.1	26.2	11.7
Average number of stocks	757	757	630	630

Table 6: Order imbalance in large orders for NYSE stocks

We run the following cross-sectional regression each month for orders greater than $10,000:

$$OIB_{i,t} = a_t + \sum_{j=1}^{12} b_{t,j} R_{i,t-j} + c_t \ln\left(Size_{i,t-1}\right) + d_t \ln\left(BM_{i,t-1}\right) + e_t Vol_{i,t-1} + \sum_{j=1}^{3} f_{t,j} OIB_{i,t-j} + u_{i,t}.$$

The table reports the time-series averages of the coefficients together with their Newey-West corrected *t*-statistics within parentheses. OIBNUM is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. Return is in percent per month, Size is market capitalization in millions of dollars, BM is book-to-market, and Vol is the standard deviation of returns in percent per month calculated using daily returns within the month. The sample includes all NYSE stocks over the period 1993 to 2010.

	OIBNUM	OIBSH
Constant	-7.18 (-5.82)	-14.66 (-7.41)
Return(-1)	-8.38 (-9.87)	-1.19 (-1.76)
Return(-2)	-4.86 (-8.74)	-2.15 (-3.04)
Return(-3)	-2.42 (-5.26)	-0.71 (-1.07)
Return(-4)	0.11 (0.28)	0.68 (1.24)
Return(-5)	1.32 (3.77)	1.56 (3.06)
Return(-6)	1.38 (3.17)	2.00 (3.23)
Return(-7)	1.04 (2.38)	1.23 (2.55)
Return(-8)	1.13 (2.93)	0.01 (0.03)
Return(-9)	1.23 (2.95)	1.48 (2.72)
Return(-10)	0.76 (1.89)	0.63 (1.28)
Return(-11)	1.03 (2.46)	0.86 (1.73)
Return(-12)	1.49 (3.64)	1.33 (2.55)
ln(Size(-1))	0.56 (7.18)	1.18 (8.87)
ln(BM(-1))	-0.61 (-9.09)	-0.16 (-2.32)
Vol(-1)	0.56 (0.57)	0.27 (0.24)
OIB(-1)	16.76 (16.96)	10.96 (14.39)
OIB(-2)	9.06 (13.12)	6.87 (10.67)
OIB(-3)	8.72 (13.62)	6.81 (10.81)
Average $\bar{R}^2$ (%)	11.5	7.5
Average number of stocks	1,318	1,318

Table 7: Cross-sectional determinants of order imbalance for Nasdaq stocks

We run the following cross-sectional regression each month:

$$OIB_{i,t} = a_t + \sum_{j=1}^{12} b_{t,j} R_{i,t-j} + c_t \ln\left(Size_{i,t-1}\right) + d_t \ln\left(BM_{i,t-1}\right) + e_t Vol_{i,t-1} + \sum_{j=1}^{3} f_{t,j} OIB_{i,t-j} + u_{i,t}.$$

The table reports the time-series averages of the coefficients (multiplied by 100) together with their Newey-West corrected *t*-statistics (using three lags) within parentheses. OIBNUM is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. Return is in percent per month, Size is market capitalization in millions of dollars, BM is book-to-market, and Vol is the standard deviation of returns in percent per month calculated using daily returns within the month. The sample includes all Nasdaq stocks over the period 1987 to 1996.

	OIBNUM	OIBSH	OIBNUM	OIBSH	OIBNUM	OIBSH
	All months		December		January	
Constant	-26.10 (-18.85)	-34.19 (-31.23)	-44.01 (-9.92)	-50.52 (-12.60)	-8.09 (-1.68)	-23.56 (-8.02)
Return(-1)	-18.58 (-12.38)	-8.62 (-8.40)	-10.83 (-6.06)	-4.67 (-2.09)	-23.27 (-9.09)	-15.82 (-6.67)
Return(-2)	-6.44 (-9.78)	-3.65 (-5.59)	0.37 (0.55)	1.10 (2.16)	-11.78 (-6.71)	-4.98 (-2.94)
Return(-3)	-2.86 (-7.78)	-2.42 (-5.27)	2.36 (1.75)	0.05 (0.02)	-7.63 (-5.64)	-4.76 (-4.94)
Return(-4)	0.71 (1.80)	0.24 (0.47)	5.81 (4.49)	0.72 (0.54)	-4.01 (-1.69)	-3.20 (-1.61)
Return(-5)	1.07 (2.72)	-0.09 (-0.17)	5.45 (3.96)	0.30 (0.27)	-4.60 (-1.96)	-5.63 (-1.45)
Return(-6)	1.31 (3.22)	0.71 (1.27)	5.87 (3.05)	2.02 (1.50)	0.07 (0.07)	0.88 (0.40)
Return(-7)	1.63 (3.53)	1.33 (2.28)	5.98 (6.36)	4.02 (3.95)	-0.35 (-0.51)	-0.35 (-0.32)
Return(-8)	1.46 (3.10)	0.73 (1.17)	5.31 (6.33)	1.27 (1.61)	0.62 (0.87)	2.11 (2.42)
Return(-9)	1.22 (2.96)	0.74 (1.27)	4.80 (12.76)	2.82 (2.02)	-0.84 (-1.03)	0.96 (1.33)
Return(-10)	1.88 (4.61)	0.38 (0.71)	5.61 (2.34)	3.09 (1.31)	3.23 (2.90)	2.82 (5.81)
Return(-11)	1.36 (3.23)	0.06 (0.13)	3.14 (3.12)	0.41 (0.20)	0.26 (0.13)	1.43 (0.63)
Return(-12)	1.37 (2.52)	0.60 (1.03)	7.37 (8.24)	0.20 (0.10)	-1.29 (-0.65)	-2.04 (-0.70)
ln(Size(-1))	1.79 (16.71)	2.11 (29.06)	3.25 (8.37)	3.49 (12.29)	0.50 (1.36)	1.38 (9.05)
ln(BM(-1))	-1.20 (-10.14)	-0.57 (-4.08)	-1.40 (-5.81)	-0.30 (-1.30)	-0.54 (-1.34)	0.01 (0.02)
Volatility(-1)	8.89 (8.73)	12.36 (11.76)	-6.82 (-1.29)	4.67 (1.02)	14.83 (14.36)	18.82 (5.49)
OIB(-1)	28.59 (15.88)	15.99 (17.07)	28.75 (5.39)	15.40 (5.76)	25.64 (8.30)	17.26 (8.03)
OIB(-2)	12.07 (22.06)	8.98 (13.80)	11.72 (9.29)	8.09 (14.15)	12.76 (6.13)	8.90 (4.62)
OIB(-3)	10.07 (18.65)	8.28 (13.04)	8.98 (8.97)	8.09 (4.93)	12.10 (6.34)	9.49 (5.75)
Average $\bar{R}^2$ (%)	22.2	8.6	29.2	11.9	18.5	8.5
Average number of stocks	2,279	2,279	2,435	2,435	2,235	2,235

Table A1: Cross-sectional determinants of order imbalance (with no lags of order imbalance) for NYSE stocks

Panel A reports the results of the following cross-sectional regression each month:

$$OIB_{i,t} = a_t + \sum_{j=1}^{12} b_{t,j} R_{i,t-j} + c_t \ln\left(Size_{i,t-1}\right) + d_t \ln\left(BM_{i,t-1}\right) + e_t Vol_{i,t-1} + u_{i,t},$$

The table reports the time-series averages of the coefficients together with their Newey-West corrected *t*-statistics (using three lags) within parentheses. OIBNUM is the number of buys less number of sells in a month as a fraction of the total trades in that month. OIBSH is the number of shares bought less number of shares sold in a month as a fraction of the total shares traded that month. Return is in percent per month, Size is market capitalization in millions of dollars, BM is book-to-market, and Vol is the standard deviation of returns in percent per month calculated using daily returns within the month. The sample includes all NYSE stocks over the period 1993 to 2010.

	OIBNUM	OIBSH
Cnst	-6.96 (-3.75)	-19.71 (-9.39)
Return (-1)	-7.79 (-5.83)	-0.47 (-0.90)
Return(-2)	-4.87 (-5.65)	-1.07 (-2.40)
Return(-3)	-2.59 (-4.38)	0.43 (0.90)
Return(-4)	-1.14 (-2.67)	0.57 (1.35)
Return(-5)	0.67 (1.78)	1.29 (3.30)
Return(-6)	1.42 (3.75)	1.73 (4.00)
Return(-7)	1.79 (4.92)	1.24 (3.51)
Return(-8)	2.21 (5.27)	1.15 (3.16)
Return(-9)	1.80 (3.82)	1.30 (3.41)
Return(-10)	1.97 (4.27)	0.70 (2.31)
Return(-11)	2.36 (5.22)	0.59 (1.81)
Return(-12)	2.72 (5.93)	1.52 (4.58)
ln(Size(-1))	0.54 (5.53)	1.55 (11.20)
ln(BM(-1))	-1.52 (-8.70)	-0.27 (-4.49)
Vol(-1)	4.22 (2.87)	4.87 (4.85)
Average $\bar{R}^2$ (%)	7.5	6.0