Distribution Agreement

Signature:

__________________ ____________________________

Qing Tong Date

Three Essays in Financial Economics

By
Qing Tong
Doctor of Philosophy

Business

Jay A. Shanken, Ph.D.
Advisor

Tarun Chordia, Ph.D.
Co-advisor

Jeffrey A. Busse, Ph.D.
Committee Member

T. Clifton Green, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Three Essays in Financial Economics

By

Qing Tong
M.Sc., Virginia Tech, 2003

Advisor: Tarun Chordia, Ph.D., UCLA, 1993
Co-advisor: Jay A. Shanken, Ph.D., Carnegie Mellon University, 1983

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Business
2010

Abstract

Three Essays in Financial Economics

By Qing Tong

In the first essay, ("Abnormal Volume in Large Trades and the Cross-Section of Expected Stock Returns"), I employ a new variable, abnormal volume in large trades, to study information risk. I provide new evidence to support the asymmetric information hypothesis that stocks with greater information risk are expected to have higher returns. In the second essay, ("Retail Investor Industry Herding", joint work with Russell Jame), we examine the industry wide investment decisions of individuals (retail investors). We find that retail investor herd into industries, and that industry herding can forecast industry returns. The industries most heavily bought by retail investors significantly underperform the industries most heavily sold by retail investors over the subsequent 3 to 12 months. In the third essay, ("Mutual Fund Industry Selection and Persistence", joint work with Jeff Busse), we analyze mutual fund industry selectivity—the performance of a fund's industry allocation relative to the market. We find that industry selection accounts for a quarter of fund performance based on two-digit SIC codes, with the remaining attributable to the performance of individual stocks relative to their own industries. We find that industry-selection skill drives persistence in relative performance, particularly over longer investment horizons. Unlike individual-stock-selection ability, industry selectivity is not eroded by increasing fund assets.

Three Essays in Financial Economics

By

Qing Tong
M.Sc., Virginia Tech, 2003

Advisor: Tarun Chordia, Ph.D., UCLA, 1993
Co-advisor: Jay A. Shanken, Ph.D., Carnegie Mellon University, 1983

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Business
2010

Acknowledgements

I express my deepest gratitude to my advisors, Jay Shanken and Tarun Chordia, for their support, guidance, and wisdom during my study here at Emory University. I would like to thank Jeff Busse, Clifton Green, Byoung Hwang, Russell Jame, Narasimhan Jegadeesh, Yue Tang for many helpful comments. I would also like to thank my parents, Jiaming Tong and Zhongli Yun for their encouragement throughout my scholarly pursuits. Lastly, I would like to thank my wife, Wei Zhang who has been by my side through all of this. Without her unconditional love and unimaginable patience, my thesis work would not have been possible.

Table of Contents

List of Tables

List of Figures

Introduction

This dissertation consists of three essays. In the first essay ("Abnormal Volume in Large Trades and the Cross-Section of Expected Stock Returns"), I employ a new variable, abnormal volume in large trades, to measure the intensity of information-based trading. I find that abnormal volume in large trades is positively priced in the cross-section, after controlling for firm size, book-to-market ratio, momentum, share turnover, the volatility of turnover, the Amihud illiquidity measure and PIN. This effect survives a number of robustness checks and is statistically and economically significant. My finding provides new evidence to support the asymmetric information hypothesis that stocks with greater information risk are expected to have higher returns.

In the second essay ("Retail Investor Industry Herding", joint work with Russell Jame), we examine the industry-wide investment decisions of individuals (retail investors). We find strong evidence that retail investors herd into and out of the same industries. Retail investor industry herding is distinct from firm-level herding and persists even after controlling for herding into stocks with similar size and book-to-market ratios. Moreover, retail investor industry herding forecasts industry returns. Over weekly horizons, industries heavily bought by retail investors significantly outperform industries heavily sold by retail investors, while over quarterly horizons industries heavily bought by retail investors significantly underperform industries heavily sold by retail investors. We decompose the poor performance of retail trades documented by Barber, Odean, and Zhu (2008) and Hvidkjaer (2008) and estimate that roughly 60% is due to poor industry selection. Taken together, our results suggest that retail investors categorize stocks by industry and that industry-wide sentiment contributes significantly to the poor performance of retail investors.

In the third essay ("Mutual Fund Industry Selection and Persistence", joint work with Jeff Busse), we analyze mutual fund industry selectivity—the performance of a fund's industry allocation relative to the market. We find that industry selection accounts for a quarter of fund performance based on two-digit SIC codes, with the remaining attributable to the performance of individual stocks relative to their own industries. We find that industry-selection skill drives persistence in relative performance, particularly over longer investment horizons. Unlike individual-stock-selection ability, industry selectivity is not eroded by increasing fund assets.

First Essay: Abnormal Volume in Large Trades and the Cross-section of Expected Stock Returns

1. Introduction

An influential set of recent papers by Easley, Hvidkjaer, and O'Hara (2002) and Easley and O'Hara (2004) argue that asymmetric information should affect the cross-section of asset expected returns. When information is private and uninformed investors cannot perfectly infer it from prices, the uninformed view the asset as being risky. Private information creates a risk for uninformed traders as the trading gains of the informed arise from the trading losses of the uninformed. Less informed traders recognize they are at an information disadvantage and will try to hold assets where their disadvantage is less. Easley and O'Hara construct a rational expectations equilibrium asset pricing model with asymmetric information and show that, holding all else constant, uninformed investors demand a premium to hold shares in firms with higher information asymmetry.[1]

It is difficult to test whether asymmetric information should be priced because the extent of private information is not directly observable. Easley, Hvidkjaer, and O'Hara (2002) use a structural market microstructure model to generate a measure for the probability of information-based trading (PIN) in individual stocks. PIN is the ratio of the arrival rates of information-induced orders and all orders, which includes both informed and uninformed orders. They estimate this measure using high-frequency data for NYSE-listed stocks for the period 1983 to 1998. The resulting estimates are a time series of individual stock probabilities of information-based trading for a large cross section of stocks. They show that PIN is positively and significantly related to average stock returns.

[1] In Easley and O'Hara's model, the number of assets is finite. Hughes, Liu, and Liu (2007) and Lambert, Leuz, and Verrecchia (2007) argue that the effect of asymmetric information on expected returns is diversifiable in a large economy. They show that Easley and O'Hara's argument does not hold in a large economy because prices will eliminate asymmetry in information by fully revealing private signals.

However, the use of PIN to test the asymmetric information hypothesis has been questioned recently. Duarte and Young (2009) show that PIN is priced not because of information asymmetry, but because of a liquidity effect that is unrelated to asymmetric information. Aktas et al. (2007) document that the PIN variable decreases before announcements of M&A transactions and increases after the announcement, which is counter-intuitive because there is considerable evidence of information leakage prior to such M&A announcements.[2] In addition, Aktas et al. (2007) and Aslan et al. (2007) point out a convergence problem in the PIN estimation. They state that for stocks with a very large number of trades, the optimization program encounters computational underflow and is unable to evaluate the likelihood function. This problem would be aggravated for more recent samples since trading frequency has increased dramatically.

The goal of this paper is to develop a new measure to proxy for informed trading and to test the asymmetric information hypothesis. I employ a new methodology to separate trades into different categories and propose a simple measure, *abnormal volume in large trades* (ALT), to capture information-based trading. Many theoretical and empirical studies have shown that informed traders prefer to trade in large size and large trades have information content.[3] For example, Easley and O'Hara (1987) argue that risk-averse informed traders prefer to trade at the known large-quantity price rather than trading at uncertain prices with a multiple-small-trade strategy. Holden and Subrahmanyam (1992) show that when multiple informed traders exist, an informed trader will trade very aggressively to compete with other informed traders. Hasbrouck (1988, 1991) find the price impact of trades on subsequent quotes to increase with trade size.

---

[2] See, for example, Dennis and McConnell (1986), Keown and Pinkerton (1981) and Meulbroek (1992).

[3] Informed traders may hide their trading activity by submitting medium-size orders. Empirically, Barclay and Warner (1993) find that the price movements are mainly due to medium-size trade category with trade size from 1,000 to 9,990 shares. Those trades account for 45.7% of the total trades but 92.8% of the cumulative price change. Large trades account for 1.7% of trades and 9.5% of price change. The Large-trade criterion in my paper is quite different from their paper. The average cutoff for large trade across my entire sample is about 3,000 shares (60,000 dollar volume). Therefore large trades in this paper overlap the medium-size category in Barclay and Warner (1993).

Hwang and Qian (2009) show that large trades have a significantly larger permanent price effect than small trades.[4] [5]

Therefore, the occurrence of a great amount of unexpected large trades may imply that informed trading has increased. Also, I use the change of large trades rather than the level of large trades because the change is less likely to relate to firms' other characteristics such as size and liquidity.

My main results show that ALT is positively priced in the cross-section of expected stock returns, even after controlling for other known determinants of expected returns such as firm size, book-to-market ratio, momentum, share turnover, the volatility of turnover, and the Amihud liquidity measure. My finding is robust to the Fama and French (1993) risk factor controls as well as to the estimation of factor loadings conditional on macroeconomic variables and firm characteristics such as size and book-to-market ratio. The pricing of ALT is also economically significant. A one standard deviation variation of ALT results in an annual premium that ranges from 2.3% to 2.9%.

Is ALT a better return predictor than PIN? I run a "horse race" to compare ALT with PIN. I find that ALT has a more robust relationship with expected stock returns than PIN. Specifically, PIN is not significant when including either the Amihud illiquidity measure, momentum, PSOS (a component of PIN that is related to illiquidity),[6] or ALT as a control

---

[4] In contemporaneous work, Hwang and Qian (2009) develop a different information risk measure and show that information risk is priced. I found their work after the first version of my paper was written. Their measure is based on the price discovery of large trades, estimated via a vector error-correction model. Both of our measures are built on the notion that large trades are more likely to be associated with informed trading.

[5] Many other papers also suggest that large trades have information content. See Stoll (1978), Pfleiderer (1984), Glosten and Harris (1988), Hausman, Lo, and MacKinlay (1992), Holthausen and Verrecchia (1990), Seppi (1990), Kim and Verrecchia (1991), Huang and Stoll (1997), Easley, Kiefer, and O'Hara (1997) and Koshi and Michaely (2000).

[6] Duarte and Young (2009) decompose PIN into two components, adjPIN (related to asymmetric information) and PSOS (related to illiquidity). See Duarte and Young (2008) for a detailed description. I am very grateful to Lance Young for providing their data.

variable. ALT remains significantly positive when I use the above mentioned control variables in cross-sectional regressions.

I also investigate the relation between ALT and changes in small trades and medium trades and examine whether they have different predictive ability for stock returns. I find that ALT has very low correlations with the change in small trades, which suggests that ALT is a measure of informed trading, rather than just a proxy for public information. Otherwise we would see high correlation because the occurrence of small trades has been documented to be related to public information.[7] Unlike ALT, changes in small trades and medium trades have little or no relationship with future returns. This shows that large trades vis-a-vis other types of trades have different information content. In addition, I include the change in total trading volume as a regressor in the cross-sectional regression. I find that the coefficient on abnormal trading volume is not significant after controlling for ALT while the coefficient on ALT remains positive and significant. This indicates that my results are not likely to be driven by the visibility hypothesis of Gervais, Kaniel, and Mingelgrin (2001).[8]

Next, I take an extra step to present some additional evidence that the occurrence of large trades is correlated with information-based trading.

I investigate the pattern of large trade volume prior to important corporate announcements, when the amount of asymmetric information is likely to increase. I consider both unscheduled and scheduled corporate announcements. Scheduled (unscheduled) corporate

---

[7] For example, Odean (1998) and Barber and Odean (2008) show that retail investors are buyers of attention-grabbing stocks.

[8] Gervais, Kaniel, and Mingelgrin (2001) document the intertemporal role of abnormal trading volume in predicting directional price changes. Specifically, Gervais et al. (2001) find that stocks experiencing unusually high (low) trading volume over a day or a week tend to appreciate (depreciate) over the course of the following month. They argue that this evidence is consistent with the visibility hypothesis, suggested by Miller (1977) and Mayshar (1983). By Miller (1977) and Mayshar (1983), if traders have diverse opinions about the value of a stock, the traders who end up holding that stock will be the most optimistic about its value. They further argue that if the stock's supply is limited because of constraints on short-selling, the opinions of the pessimistic traders will fail to be fully incorporated into the stock's price. As a consequence, any shock that attracts the attention of investors towards a given stock should result in a subsequent demand and price increase, as the set of potential buyers then includes a larger fraction of the market, whereas the set of potential sellers is largely restricted to the current stockholders.

announcements refer to the public availability (unavailability) of when an announcement will be issued. Following Chae (2005), I use mergers announcements as unscheduled announcements and earnings announcements as scheduled announcements. As noted by Chae (2005), these two types of announcements are chosen because they represent major corporate events that have substantial impacts on stock prices.[9]

I find that during the period (ten to four days) prior to earnings announcements, the amount of volume for each type of trade decreases. The decreased trading volume is probably due to reduced liquidity trading before scheduled announcements.[10] Although the amount of large trades decreases before earnings announcements, the magnitude of decrease of large trades is lower than that of other types of trades. When the time approaches closer to earnings announcements (three to one day prior to earnings announcements), information leakage might increase. I find that the volume for each type of trade increases and the increase in large trades is the highest.

In the case of merges announcements (unscheduled announcements), uninformed investors cannot predict informed trading patterns. Liquidity trading is kept relatively unchanged compared to that before scheduled announcements. Therefore, the increased trading volume is more likely to reflect the intensity of private information. I find that during both periods (ten to four days and three to one day) before mergers announcements, the increase in large trade is higher than the increase in other types of trades.

I also find that during the period (three to ten days) after earnings and mergers announcements, the increase in small trades is the highest while the increase in large trades is the lowest for all types of trades. In sum, these trading patterns around corporate events are consistent with the notion that large trades are more likely to be associated with informed trading.

[9] See Ball and Brown (1968), Dennis and McConnell (1986), Bernard and Thomas (1990), Vega (2006) and many others for studies on stock prices around such events.

[10] See Admati and Pfleiderer (1988) and Foster and Viswanathan (1990).

Overall, my main contributions are as follows. I employ ALT to measure information-based trading and find that ALT affects stock price in the cross-section. My findings are consistent with the asymmetric information hypothesis that stocks with greater information risk are expected to have higher returns. My results also show that ALT is a better return predictor than PIN. Additionally, as an extension of Chae (2005), where he studies the changes in total trading volume around scheduled and unscheduled announcements, I investigate the trading patterns for different types of trades.

The remainder of the paper is outlined as follows. Section 2 presents the methodology. Section 3 describes the data. Section 4 shows that ALT is priced in the cross-section of expected stock returns. Section 5 compares ALT with PIN. Section 6 provides additional evidence that large trades are associated with informed trading. Section 7 concludes the paper.

2. Methodology

2.1 Decomposition of trading volume

Simple cutoffs have been employed for trade classification in the literature. For example, Barber, Odean and Zhu (2008) and Frazzini and Lamont (2007) define a trade as a large trade (small trade) if the dollar trading volume is greater than $50,000 (less than $5,000). Bessembinder and Kaufman (1997) classify the trades into three categories: small trade (less than $10,000), medium trade ($10,000-$199,999), and large trade (above $200,000). Hvidkjaer (2006) suggests sorting stocks into quintiles and using different cut-off points conditioning on firm size.[11] However, as noted by Barber et al. (2008) and Hvidkjaer (2008), changes in market environment (such as decimalization in 2001) have caused simple classifications to be less reliable after 2001. Chordia, Roll, and Subrahmanyam (2008) also document that the average trade size is changing over time.

[11] For example, Hvidkjaer (2006) suggests that within the smallest size quintile firms, cut-off points for large trade (small trade) are $6,800 ($3,400), while within the largest size quintile firms, cut-off points for large trade (small trade) are $32,800 ($16,400).

This paper provides a new methodology for trade classification. The procedure used to create my classification is a dynamic process, which is based on each individual stock's trades in the past year. The method is as follows:

(a) At the beginning of each month t, sort each stock's past one-year intraday transactions (say, one million transactions) from the lowest to the highest dollar trading volume and then find the 30%, 60%, and 90% fractiles. I classify trades into four categories: (i) small trade: less than the 30% fractile; (ii) medium trade size1: between the 30% fractile value and the 60% fractile; (iii) medium trade size2: between the 60% fractile value and the 90% fractile; (iv) large trade: greater than the 90% fractile. For example, based on the transactions from January 1995 to December 1995, the 30%, 60% and 90% fractiles for General Electric Co. (ticker GE) are $5,212, $70,750 and $116,875, respectively. I sum up dollar trading volume for each type of trade across the year.

(b) In month t, each stock trade's type is determined by comparison to the stock trade's 30%, 60% and 90% fractile value in the past one year. For example, in January 1996, the trade classification for GE is based on the transactions from January 1995 to December 1995. A trade for GE is defined as a large trade if the trade is greater than $116,875, and a trade is defined as a small trade if the trade is less than $5,212. I sum up dollar trading volume for each type of trade across January 1996.

The process is repeated every month. For example, in February 1996, the trade classification for GE is based on the past one year GE transactions from February 1995 to January 1996.

(c) Abnormal total trading volume (AT) in month t is defined as

$$\text{AT(t)}=\frac{\text{total dollar trading volume in month t}}{\frac{1}{12}\text{total dollar trading volume in past one year}}-1 \qquad (1)$$

Abnormal small trading volume (AS) in month t is defined as

$$\text{AS(t)}=\frac{\text{summed small trades dollar volume in month t}}{\frac{1}{12}\text{summed small trades dollar volume in past one year}}-1 \quad (2)$$

Abnormal medium-size1 trading volume (AMT1) in month t is defined as

$$\text{AMT1(t)}=\frac{\text{summed medium-size1 trades dollar volume in month t}}{\frac{1}{12}\text{summed medium-size1 trades dollar volume in past one year}}-1 \quad (3)$$

Abnormal medium-size2 trading volume (AMT2) in month t is defined as

$$\text{AMT2(t)}=\frac{\text{summed medium-size2 trades dollar volume in month t}}{\frac{1}{12}\text{summed medium-size2 trades dollar volume in past one year}}-1 \quad (4)$$

Abnormal large trading volume (ALT) in month t is defined as

$$\text{ALT(t)}=\frac{\text{summed large trades dollar volume in month t}}{\frac{1}{12}\text{summed large trades dollar volume in past one year}}-1 \quad (5)$$

2.2 Cross-sectional regressions

The main test adopts the methodology by Brennan, Chordia, and Subrahmanyam (1998) and Avramov and Chordia (2006), who test factor models by regressing risk-adjusted returns on firm-level attributes such as size, book-to-market, and turnover. The use of single securities in empirical tests of asset pricing models guards against the data-snooping biases inherent in portfolio based asset pricing tests (Lo and MacKinlay (1990)) and avoids the loss of information that results when stocks are sorted into portfolios (Litzenberger and Ramaswamy (1979)).

More specifically, I first regress the excess return on stock j, (j=1,..,N) on asset pricing factors, F_{kt}, (k=1,..,K) allowing the factor loadings, β_{jk}, to vary over time as function of stock-firm size and book-to-market ratio as well as macroeconomic variables.[12] The conditional factor loadings of security *j* are modeled as:

[12] The method of allowing betas to vary with macroeconomic variables is first employed in Shanken (1990).

$$\beta_{jk}(t-1) = \beta_{jk1} + \beta_{jk2} z_{t-1} + \beta_{jk3} Size_{jt-1} + \beta_{jk4} BM_{jt-1} \qquad (6)$$

where $Size_{jt-1}$ and BM_{jt-1} are the market capitalization and the book-to-market ratio at time $t-1$, and z_{t-1} denotes a vector of macroeconomic variables: the term spread, the default spread and the T-bill yield. The term spread is the yield differential between Treasury bonds with more than ten years to maturity and T-bills that mature in three months. The default spread is the yield differential between bonds rated BAA and AAA by Moodys. In the empirical analysis the factor loadings, $\beta_{jk}(t-1)$ are modeled using three different specifications: (i) the unconditional specification in which $\beta_{jkl} = 0$ for l>1, (ii) the firm specific variation model in which the loadings depend only on firm level characteristics, $\beta_{jk2} = 0$, and (iii) the macro $\beta_{jk3} = \beta_{jk4} = 0$.

The dependence on size and book-to-market is motivated by the general equilibrium model of Gomes, Kogan, and Zhang (2003), which justifies separate roles for size and book-to-market as determinants of beta. In particular, firm size captures the component of a firm's systematic risk attributable to its growth option, and the book-to-market ratio serves as a proxy for risk of existing projects. Incorporating business-cycle variables follows the extensive evidence on time series predictability using macroeconomic variables (see, for example, Keim and Stambaugh (1986), Fama and French (1989) and Chen (1991)).

I then run cross-sectional regressions of risk-adjusted returns, rather than gross returns, as dependent variables on the equity characteristics:

$$R_{jt} - R_{Ft} - \sum_{k=1}^{K} \hat{\beta}_{jkt-1} F_{jk} = c_{ot} + \sum_{m=1}^{M} c_{mt} Z_{mjt-2} + e_{jt} \qquad (7)$$

where $\hat{\beta}_{jkt}$ is the conditional beta estimated by a first-pass time-series regression over the entire sample period.[13] Z_{mjt-2} is the value of characteristic m for security j at time t-2,[14] and M is the

[13] Avramov and Chordia (2006) have shown that using the entire time series to compute the factor loadings gives the similar results as using rolling regressions.

[14] All characteristics were lagged by two months to avoid biases because of bid-ask effects and thin trading (see Jegadeesh (1990) and Brennan, Chordia, and Subrahmanyam (1998)).

total number of characteristics. The standard Fama-MacBeth (1973) estimators are the time-series averages of these coefficients, $\hat{c}_t$. The standard errors of the estimators are traditionally obtained from the time series of monthly estimates. I correct the Fama-MacBeth (1973) standard errors, attributable to the error in the estimation of factor loadings in the first-pass regression, using the approaches in Shanken (1992).

The firm characteristics included are (i) SIZE: measured as the natural logarithm of the market value of equity, (ii) BM: logarithm of the book to market ratio, (iii) TURN: the logarithm of the ratio of monthly share trading volume and shares outstanding, (iv) RET2-12: the cumulative return over the eleven months ending at the beginning of the previous month, (v) CVTURN: the natural logarithm of the coefficient of variation of turnover calculated over the past 36 months,[15] and (vi) AT, ALT, AMT1, AMT2 and AST.

Under the null hypothesis of exact pricing, all these characteristics should be insignificant in the cross sectional regressions. Significant coefficients would point to the inefficacy of models. Brennan, Chordia, and Subrahmnyam (1998) and Avramov and Chordia (2006) find that predictive ability of size, book-to-market, turnover, and past returns is unexplained by most models. Here, I explore whether AT, ALT, AMT1, AMT2 and AST capture elements of expected returns that are not captured by the factor pricing model.

3. Data

3.1 Data description

The sample includes NYSE and AMEX stocks in the period January 1983 through December 2006. Certificates, American depository receipts, shares of beneficial interest, units, companies incorporated outside the U.S., Americus Trust components, closed-end funds, and real estate investment trusts are eliminated from the sample. Stocks included in the monthly analysis satisfy the following criteria: (i) return in the current month and over the past 36 months is

---

[15] This variable is first used in Chordia, Subrahmanyam, and Anshuman (2001) as a proxy for the volatility of liquidity.

available from CRSP, and sufficient data are available to calculate the size and turnover, (ii) sufficient data are available on the Compustat tapes to calculate the book-market ratio as of December of the previous year. To avoid extremely illiquid stocks, I eliminate penny stocks, that is, stocks with prices less than one dollar, from the sample. This screening process yields an average of 2,537 stocks per month.

Transactions data are obtained from the Institute for the Study of Security Markets (ISSM) (1983-1992) and the Trade and Quote (TAQ) data sets (1993-2006).

I/B/E/S data from 1983 to 2006 are used for the earnings announcement sample. The reporting dates are extracted from the I/B/E/S summary files. The total number of earnings announcements in the sample is 169,083.

Acquisition and target announcements for NYSE and AMEX stocks are collected from the SDC database complied by Thomson Financial Securities Data from 1983 to 2006. SDC's merger and acquisition database provides 17,298 target announcements and 42,867 acquisition announcements.

3.2 Summary statistics

Table 1 presents summary statistics. Panel A shows the average percentage of number of trades and dollar trading volume for each trade category. For example, the average percentage of number of large trade is 9.7%. The average percentage of dollar trading volume of large trade is 33.2%, indicating that relatively few numbers of large trades are associated with 1/3 of total trading dollar volume.

Panel B in Table 1 presents the summary statistics for changes in total trades (AT), in small trades (AST), in medium trades (AMT1 and AMT2) and in large trades (ALT). AT is 0.068, showing that total trading volume on average has increased over time.[16] Among the

[16] See Chordia, Roll, and Subrahmanyam (2009) for detailed discussions of the reasons that trading volume has increased.

average change in different types of trades, AST has the highest number (0.086). By contrast, the average change in large volume (ALT) has the lowest increase (0.066).

Figure 1 shows the pattern for average dollar volume cutoff of large trades across stocks over time. The average cutoff for large trade is about 60,000 dollar volume (3,000 shares).[17] It is clear to see that cutoffs have declined over time.

3.3 Correlations

Table 2 presents the time-series average of cross-sectional correlations between AT, ALT, AMT1, AMT2 and AS, as well as a number of other variables employed in this study. Correlations between AT, ALT, AMT1, AMT2 and AST are positive, which suggests an increase (or decrease) in one category of trading volume is associated with another. However, the correlations are not high. The time-series means of correlation between ALT and AT is 0.281. Also ALT is weakly correlated with AST, with a positive correlation 0.045. These correlations suggest that changes in large trades do not necessarily represent changes in other types of trades.

Table 2 also indicates that ALT has low correlations with liquidity measures. The correlations between ALT and share turnover, the Amihud illiquidity measure[18] and the quoted spread[19] are 0.283, -0.043 and 0.071, respectively. From Table 2, we can also see that ALT is positively correlated with PIN, but the correlation is low. By contrast, the correlation between AST and PIN is negative.

4. Main results

[17] The average stock price is about 20 dollars.

18 The Amihud measure is calculated as the monthly average of the ratio of the daily absolute return to daily total volume,

$$\mathrm{ILLIQ}_{it} = \frac{1}{D_{it}} \sum_{t=1}^{D_{it}} \frac{|R_{itd}|}{DVOL_{itd}} * 10^6,$$

where R_{itd} is the daily return, $DVOL_{itd}$ is the dollar trading volume of stock i on day d in month t and D_{it} is the number of days in month t for which data is available for stock i.

[19] The quoted spread is measured as the average of all quoted spread observations for each stock throughout a given month. In the literature, spread has been used as a proxy for illiquidity (see, for example, Amihud and Mendelson (1986)). It has also been used as a measure of asymmetric information (see, for example, Llorente et al. (2002)).

4.1 Portfolio sorts

Before moving on to the regression analysis, I report mean returns for the 5 portfolios formed by sorting the component stocks into quintiles each month according to the changes in trading volume at month t-1. Gross returns as well as the market model (CAPM) and Fama and French (1993) (FF) intercepts (alphas) for these portfolios are presented in Table 3. I find that excess returns and alphas increase monotonically with the ALT quintile. The differences in gross returns and alphas between the extreme quintiles are all statistically significant for the ALT portfolios. The magnitude of the differences is also economically significant. For example, the FF alpha of the high ALT portfolio exceeds that of the low ALT portfolio by 65 bp per month. However, for the AMT1, the AMT2, and the ATS portfolios, the differences are not statistically significant.

In order to distinguish between the effects of ALT and firm size, I sort stocks first by size and then by ALT into 25 portfolios and present the results in Panel A of Table 4. In each case, within each size quintile, the differential gross returns and Fama-French alphas are significant at the 10% level across the extreme ALT quintiles. Thus, the return differential across the portfolios sorted by ALT is not a phenomenon confined to only the smaller stocks. However, the differential between the extreme ALT quintiles is generally larger for smaller firms.

Panel B in Table 4 shows the effect of ALT on stock returns controlling for abnormal total trading volume (AT). Gervais, Kaniel, and Mingelgrin (2001) document the positive relationship between abnormal trading volume and future stock returns. Panel B examines whether the compensation for ALT is simply a manifestation of a return effect related to abnormal trading volume. The evidence points to a role for ALT over and above AT in predicting stock returns. Within each AT quintile, the differences in both gross returns and Fama-French alphas between the highest and lowest ALT portfolios are significant.

4.2 Regression results

I now present the results of cross-sectional Fama-Macbeth type regressions of risk-adjusted returns on firm characteristics. The characteristics used are those described in Section 2.2. Results are presented both for unconditional as well as for conditional factor loadings. The conditional factor loadings are allowed to depend on firm size and book/market ratio, as well as the term structure and default variables. For each of our factor model specifications, I document the time-series averages of the monthly cross-sectional regression coefficients and the associated t-statistics corrected using the procedure of Shanken (1992).

Table 5 reports results when three Fama-French (1993) factors are used to calculate the risk-adjusted returns. I have verified that the results are qualitatively the same when using the excess market return as a risk factor. The results conform to earlier findings. The book/market ratio is significant. The longer-term momentum variables are also significant, confirming the well-known momentum effect of Jegadeesh and Titman (1993). Turnover is negatively associated with risk adjusted returns. The significance of turnover is consistent with the evidence of Datar, Naik, and Radcliffe (1998) as well as Brennan, Chordia, and Subrahmanyam (1998). Consistent with Chordia, Subrahmanyam, and Anshuman (2001), CVTURN, a measure of liquidity volatility, negatively affects returns.

I also find that the coefficient on size is insignificant. The lack of a size effect may be due to two reasons. First, I only consider NYSE/AMEX stocks, and the size effect may be more prevalent in the smaller Nasdaq stocks. Second, earlier work (Brennan, Chordia, and Subrahmanyam (1998) and Fama (1998)) indicates that the size effect is not stable over time after its discovery by Banz (1981) in the early 1980s.

While AT and ALT are significant in isolation, the coefficient on ALT is twice as much as that on AT. Moreover, the significance of AT disappears including ALT, while ALT remains highly statistically significant in the presence of ALT. This indicates that predictive power of AT emanates entirely from ALT. When ALT, AMT2, AMT1 and AST are included in the regression, only ALT is significant.

To address the issue of the economic magnitude of the premium for ALT, I consider the coefficients of ALT in Table 5. The coefficients range from 0.27 to 0.33. Relating these to the summary statistics in Panel B of Table 1, I find that a one-standard deviation move (0.712) in ALT implies an annual ALT premium that ranges from 2.3% to 2.9%. Therefore, the return required as compensation for ALT is statistically and economically significant.

The use of conditional betas in calculating the risk adjusted returns has essentially no effect on the results. Overall, these findings show that ALT is significantly related to expected stock returns.

5. ALT and PIN

This section compares the stock return predictive ability of PIN and ALT. Before showing the results, this section briefly introduces PIN and related recent studies.

The PIN measure is based on a structural sequential trade model developed by Easley and O'Hara (1987, 1992b) and Easley et al. (1996). In the model, information events are assumed to be independent across days and to occur with probability α. When an information event occurs, it is either bad news, with probability δ, or good news with probability, $1-\delta$. On any day, independent Poisson processes determine the arrival rate of uninformed buyers and uninformed sellers. These both arrive at rate ε. On events days, the arrival rate of informed traders is μ. The probability of information-based trading is then given by

$$\text{PIN}=\frac{\alpha\mu}{\alpha\mu+2\varepsilon} \tag{8}$$

For a given stock for a certain time period, PIN is the ratio of the arrival rates of information-induced orders and all orders, which includes both informed and uninformed orders. In other words, PIN measures the proportion of all orders that might be information-induced. In order to obtain PIN, the model parameters Θ={α, μ, δ, ε}are estimated by the maximization of a likelihood function.[20]

The PIN has been widely used as an explanatory variable in studying the relationship between informed trading and a variety of issues in empirical finance such as spreads of less frequently traded stocks (Easley, Kiefer, O'Hara, and Paperman (1996)), stock splits (Easley, O'Hara, and Saar (2001)), the importance of trade size (Easley, Keifer, and O'Hara (1997)), stock analyst coverage (Easley, O'Hara, and Paperman (1998)), the role of purchased order flows (Easley, Kiefer, and O'Hara (1996)) and testing for market efficiency (Vega (2006)).

Easley, Hvidkjaer, and O'Hara (2002) employ PIN to test whether stocks with more informed trading are associated with higher expected returns. They show that PIN is positively and significantly related to average stock returns for the period from 1983 to 1998. A 10% difference in the PINs of two stocks results in a 250 basis point difference in their annual expected returns.

However, Aktas et al. (2007) document that the PIN variable decreases before announcements of M&A. This is inconsistent with the evidence of information leakages during the pre-event period and raises some concerns about the use of PIN as an information-based trading indicator. In addition, Aktas et al. (2007) and Aslan et al. (2007) point out a convergence

---

[20] Let us define (B, S) as the total number of buys and sells for a single trading day. The likelihood of observing B buys and S sells for the day is

$$L(B,S|\Theta) = \alpha\delta \exp(-\varepsilon)\frac{\varepsilon^B}{B!}\exp(-(\mu+\varepsilon))\frac{(\mu+\varepsilon)^S}{S!} + \alpha(1-\delta)\exp(-(\mu+\varepsilon))\frac{(\mu+\varepsilon)^B}{B!}\exp(-\varepsilon)\frac{(\varepsilon)^S}{S!}$$
$$+(1-\alpha)\exp(-\varepsilon)\frac{\varepsilon^B}{B!}\exp(-\varepsilon)\frac{\varepsilon^S}{S!}$$

The model assumes that each day the arrivals of an information event and trades, conditional on information events, are drawn from identical and independent distributions. Thus the likelihood function for T days is

$$L(M \mid \Theta) = \prod_{i=1}^{T} L(B_i, S_i \mid \Theta)$$

Maximization of the above equation with respect to the parameter vector Θ, yields maximum likelihood estimates of the parameters of interest.

problem in the PIN estimation. They state that for stocks with a very large number of trades, the optimization program encounters computational underflow and is unable to evaluate the likelihood function. This problem would be aggravated for more recent samples since trading frequency has increased dramatically.

More importantly, Duarte and Young (2009) show that the PIN is priced because it is a proxy for illiquidity that is unrelated to asymmetric information. They decompose PIN into two components, adjPIN, the component related to asymmetric information, and PSOS, the component related to illiquidity. They find that PIN is priced because of the PSOS component. Another test in their paper includes the Amihud (2002) illiquidity measure in Fama-MacBeth regressions. In the presence of the Amihud illiquidity measure, the coefficient on PIN drops substantially and is no longer significant.

I study the stock return predicting ability of PIN and ALT. I first examine the predictive power of PIN. Similar to Easley, Hvidkjaer, and O'Hara (2002), the coefficient on PIN is positive and statistically significant at the 10% level (see column (1) of Table 6). However, consistent with Duarte and Young (2009), when I include the Amihud (2002) illiquidity measure or PSOS in the regression, PIN is insignificant (see columns (2) and (3) of Table 6). Furthermore, past returns are not included in the cross-sectional regressions in Easley, Hvidkjaer, and O'Hara (2002). When I include the ret2-12 variable, the significance of PIN again disappears (see column (4) of Table 6). Finally, the significance of PIN is gone when including ALT (see column (5) of Table 6). This evidence suggests that PIN is not a robust variable to predict stock returns.

Next, I examine the robustness of the predictive power of ALT. In column (6) of Table 6, I include PIN, ILLIQ and other control variables as regressors. I find that ALT remains positive and significant. Overall, the findings show that ALT positively predicts stock returns.

6. Corporate events

This section provides some evidence that increased large trades reflect more informed trading.

Easley and O'Hara (1987) assert that the amount of private information available to traders varies at different times. Many events studies about corporate announcements indicate that a considerable amount of information is released around these announcements. These releases of information often generate large price changes.[21] It seems quite plausible that there exists severe information asymmetry between informed and uninformed investors prior to such announcements.

I investigate the abnormal volume in large trades and other types of trades around unscheduled and scheduled corporate announcements. Scheduled (unscheduled) corporate announcements refer to the public availability (unavailability) of when an announcement will be issued. Quarterly earnings announcements are often routinely scheduled. As noted by Chordia, Roll, and Subrahmanyam (2001), earnings announcements are among the best candidates for scheduled announcements involving a release of relevant pricing information. Accordingly, I use earnings announcements as scheduled announcements. Among unscheduled corporate announcements, acquisition and target announcements have well-documented effects on stock returns.

Define t=0 as the event data. My benchmark measures are computed for t=[-40,-11] and I use t=[-10,+10] as the event window.[22] The abnormal trading volume for each type of trades around the event widow is defined as follows:

$$\text{Abnomral trading volume } \omega_{i,t} = \frac{v_{i,t} - \overline{\tau}_i}{\overline{\tau}_i}, \quad t=[-10,10]$$

$$\text{where } \overline{\tau}_i = \frac{\sum_{t=-40}^{t=-11} v_{i,t}}{30} \tag{9}$$

[21] Chae (2005) reports that the absolute daily price change on earnings announcements, acquisition announcements, target announcements are about 56%, 45%, 287% higher than the average absolute price change on other days in the same months.

[22] Chae (2005) uses the same windows for his main analysis. Results are qualitatively similar when I choose t=[-40,-21] as the estimation window and t=[-20,20] as the event window.

$v_{i,t}$ refers to small-size, medium-size1, medium-size2, or large-size daily dollar trading volume for stock i at day t.[23] $\overline{\tau_i}$ is the corresponding benchmark measure. $\omega_{i,t}$ refers to the daily abnormal trading volume for each type of trades.

Table 7 reports average daily abnormal trading volume $\omega_{i,t}$ in the periods t=[-10,-4], t=[1,2] and t=[3,10]. Table 7 also reports the results for t=-3,-2,-1,0. The t-statistics are computed using the White correction for heteroskedasticity.

Let us look at the results around earnings announcements first. I find a decrease of around -1.12% for small trades in the period from t=-10 to t=-4. The decrease is consistent with theoretical predictions in the literature. For example, Black (1986) and Wang (1994) argue that if there is a higher possibility of trading with an informed counterparty, then uniformed traders will participate less in the market.[24] Similarly, Admati and Pfleiderer (1988) and Foster and Viswanathan (1990) point out that if liquidity traders have timing discretion, they will postpone trading until the announcement is made.

Large trades also decrease in the period from t=-10 to t=-4. However, as we can see, the decrease in the large trades is less than that of other types of trades. The daily difference between the decrease in large trade and the decrease in small trade is 0.70% with a t-statistic of 3.39. In the day from t=-3 to t=-1, information leakage might increase. I find that trading volume for each type of trade increases during this period. Also the difference between large trade and small trade increases. For example, the difference at t=-3 is 1.09% with t-statistic of 3.95 and the difference

[23] The classification of the trade size is defined in section 2.

[24] A necessary condition for this prediction to hold is that the uninformed investor must perceive a high level of information asymmetry. Before scheduled announcements (such as earning announcements), uninformed investors can expect trading demand from informed investors and thus avoid unnecessary trading.

is increased to 2.00% with a t-statistic of 5.09 at t=-1. Over the period from t=-10 to t=-1, the cumulative daily difference is about 10%.[25]

After earnings announcements, I find that all types of trades increase, but there is a greater increase in small trades than other types of trades. This is consistent with the argument by Admati and Pfleiderer (1988) and Foster and Viswanathan (1990) that there will be increased liquidity trading after the information asymmetry is resolved. The higher amount in small trades after announcements is also consistent with the attention-grabbing hypothesis of Lee (1992) and Franizzi and Lamont (2007) that the announcement catches small investors' attention and therefore a great amount of small investors are participating. My results suggest that small trades are more likely to be associated with uninformed investors than large trades. [26]

Next, let us move to takeover announcements. Unlike earnings announcements, mergers announcements are unscheduled and abrupt. Uninformed investors can not predict when such announcement will be made until it becomes public information. The amount of liquidity trading should be relatively unchanged before these announcements compared to that prior to unscheduled announcements. Therefore, the increased trade volume is likely related to information-based trading. Results in Table 7 show that there is a relative greater increase for large trades before mergers announcements. The difference between large and small trades increases from t=-10 to t=-1. Over this period, the cumulative daily difference is about 12% for acquiring firms and 19% for target firms. In contrast, in the period from t=3 to t=10, the increase

[25] 0.70%*7+1.09%+1.68%+2.00%=9.67%

[26] Prior empirical studies on small and large trader reactions to earning announcement have consistently suggested that large traders are more informed than small traders. For example, Lee (1992) and Hirshleifer et al. (2008) show that individual investors are net buyers after earnings announcements, no matter whether the news is good or bad. Frazzini and Lamont (2007) show that there are more net buyers in large trade size than net buyers in small trade size before earnings announcements, but the pattern is reversed after earnings announcements. Bhattacharya (2001) and Battalio and Mendenahll (2005) find that small traders rely strongly on the unsophisticated seasonal random walk model (i.e., earnings this quarter will be the same as earnings for the same fiscal quarter last year) to form their earnings expectations, while large traders do not.

in large trades is less than that of other types of trades. Again, these suggest that large trades represent informed trading.[27]

Overall, the trading patterns of different types of trades around corporate events provide some evidence that large trades are related to informed trading.

7. Conclusion

An interesting question in finance is whether uninformed investors demand higher returns to hold shares in firms with higher information asymmetry. Easley, Hvidkjaer, and O'Hara (2002) and Easley and O'Hara (2004) argue that information risk is an important determinant of expected returns. However, more recently, studies have challenged the view that asymmetric information should be priced in the cross-section.

An important issue is the empirical proxy that can be used to measure informed trading. This paper uses a new measure, abnormal volume in large trades, to proxy for information-based trading. By studying the trading patterns of large trades around corporate events, I provide some additional evidence that large trades are associated with informed trading.

I document that ALT is significantly related to expected stock returns. These findings obtain in portfolio sorts of ALT, two-way portfolio sorts of size and ALT, and also are apparent in linear regressions after controlling other well-known determinants of expected returns. Furthermore, the compensation for ALT in the cross-section of stock returns is not only statistically significant, but also economically material. I also show that ALT is a much more robust return predictor than PIN. Overall, my work provides new evidence to support the asymmetric information hypothesis.

---

[27] I also separate firms into two size groups (small firms and big firms) based on NYSE stocks median size. The trading patterns are similar for small and big firms. Generally, prior to corporate events, the differences between large and small trades is greater for small firms compared to the differences for big firms.

Second Essay: Retail Investor Industry Herding

1. Introduction

There is growing evidence that investors often group stocks into categories or "styles" based on shared commonalities. For example, Barberis, Shleifer and Wurgler (2005) find that stocks added to the S&P 500 index begin to covary more with other members of the index, and Greenwood (2008) provides similar evidence for the Nikkei 225. Similarly, Green and Hwang (2008) document that stocks that undergo stock splits experience an increase in comovement with low-priced stocks and a decrease in comovement with high-priced stocks. These results are consistent with investors categorizing stocks based on index membership and price. Another potentially important category is industry. For example, Microsoft, Google, and Yahoo are often categorized as "technology stocks", while Merck, Pfizer, and Eli Lilly are often grouped together as "pharmaceutical stocks". Moreover, industry-wide categories appear important enough to merit institutional investors offering sector oriented mutual funds such as "Vanguard Utilities" or "Fidelity Wireless Portfolio".

If investors categorize stocks by industry membership, then their investment decisions will have an industry-wide component. This implies that industry-level reallocations should occur with greater intensity than reallocations across stocks grouped randomly. There are at least two reasons to expect that these industry-level reallocations will be particularly strong amongst retail investors. First, retail investors tend to have more limited resources than institutional investors. Thus, retail investors seem more susceptible to simplifying complex investment decisions by categorizing stocks by industry. Indeed, processing information on 50 different industries is far less time consuming than processing information on thousands of different stocks. Second, prior research has found strong evidence that the trading of retail investors is systematically correlated (see e.g. Kumar and Lee (2006), and Barber, Odean, and Zhu (2009b)).[28] Thus, if retail investors do categorize stocks by industry, it seems likely that the industry-wide investment decisions of individuals will aggregate into large industry-wide demand shocks.

In this paper, we explore three main questions about retail investor industry herding. First, do retail investors herd across industries? Second, how does retail investor industry herding impact industry-level prices? Third, to what extent is the poor performance of retail investor trading driven by their industry-wide investment decisions?

To answer these questions, we calculate the proportion of all trades in an industry that are buys (industry proportion bought) using the Trade and Quotes (TAQ) and Institute for the Study of Security Markets (ISSM) transaction data over the period of 1983-2000. We find strong and persistent herding by small traders (hereafter retail investors) at the industry-level. For example, the cross-sectional correlation between small trade proportion bought in week t and week $t+1$ averages over 60%. Moreover, retail investor industry herding is highly persistent. The cross-sectional correlation between industry-level proportion bought at week t and week t+52 averages 16%. In addition, we show that industry herding is distinct from firm-level herding and persists even after controlling for herding into stocks with similar market capitalizations and book-to-market ratios. Consistent with the style investing model of Barberis and Shleifer (2003), we find that retail investors tend to chase industries that have performed well over the past two years. In fact, prior industry returns can forecast retail investor firm-level proportion bought, even after controlling for prior firm-level returns.

Next, we examine the impact of retail investor industry herding on industry prices. The Barberis and Shleifer (2003) style investing model posits that style-level demand is not entirely

[28] Prior research has also found that institutional investor trading is correlated; however the magnitude of retail investor herding is generally much larger than institutional herding. For example, Lakonishok, Shleifer, and Vishny (1992) report a herding measure of 2.7% amongst pension funds and Grinblatt, Titman, and Wermers (1995) report a herding measure of 2.5% amongst mutual funds managers. In constrast, Barber, Odean, and Zhu (2008b) find that herding ranges from 6.8% amongst retail investors at a discount brokerage and 12.8% amongst retail investors at a full service brokerage.

driven by fundamentals. The model predicts that style-level demand will push prices away from fundamentals in the short run and lead to long-term reversals. However, other models theorize that herding is driven by investors receiving correlated signals about fundamentals. (e.g Froot, Scharfstein, and Stein (1992) or Hirshleifer, Subrahmanyam, and Titman (1994)).[29] These models argue that herding simply reflects the process in which value-relevant information is impounded into prices. Thus these models do not predict long-run reversals.

We find that weekly retail investor industry proportion bought positively forecasts industry returns over the subsequent week. We also find that retail investor industry proportion bought over the prior quarter (6 months or year) negatively forecasts industry returns over the subsequent quarter (6 months or year). A portfolio that went short the value-weighted quintile of industries most heavily bought over the prior quarter and went long the value-weighted quintile of industries most heavily sold would earn an average five-factor alpha of 41 basis points per month over the subsequent quarter. These results support the style investing model of Barberis and Shleifer (2003) and are inconsistent with rational explanations of industry herding.

Barber, Odean, and Zhu (2009) and Hvidkjaer (2008) find that small trade proportion bought also forecasts firm-level returns. Stocks heavily bought by retail investors, measured over the past year, significantly underperform stocks heavily sold by retail investors. To assess the extent to which the poor performance of retail investor trading is driven by industry-wide sentiment, we decompose retail investor performance into a firm-specific component and an industry-wide component. Our results indicate that industry selection is responsible for roughly 60% of the poor performance documented by Barber, Odean, and Zhu (2009) and Hvidkjaer (2008). Moreover, after controlling for industry selection, we find that the stock picking ability of retail investors is not significantly different from zero. The results suggest that industry sentiment explains a significant portion of the poor performance of retail trades.

[29] These models were designed to explain herding into specific stocks, not industries. However, it is equally plausible that investors can receive correlated signals about value-relevant industry information.

Lastly, we compare our findings of small trade industry herding with the results based on large trade ("institutional") industry herding. Consistent with prior work on institutional industry herding (e.g. Choi and Sias (2008) and Froot and Teo (2008)), we find statistically significant evidence of industry herding by institutions. However, the magnitude of institutional industry herding is roughly half the magnitude of retail investor industry herding. Moreover, we find no significant relationship between institutional industry proportion bought and longer-horizon industry returns.

This paper contributes to the growing empirical literature on style investing. To our knowledge, this is the first paper that examines the industry-wide investment decisions of retail investors. Kumar (2009) finds that retail investors herd into similar size and book-to-market styles and finds some evidence of style-level momentum. We show that even after controlling for size and book-to-market, retail investors herd at the industry level. Moreover, we are able to document both style-level momentum at weekly horizons, and style-level reversals at quarterly to yearly horizons. Choi and Sias (2008) and Froot and Teo (2008) examine industry herding, but focus exclusively on institutional investors. We show that relative to institutions, retail investors exhibit significantly greater industry herding and have a substantially different impact on industry prices. Our comparison suggests that industry herding by retail investors is more motivated by sentiment, while institutional industry herding is more motivated by informational reasons.

This paper also adds to the literature that investigates the relationship between investor sentiment and subsequent returns. Baker and Wurgler (2006) show that when economy wide sentiment is high, subsequent returns for stocks that are difficult to value (i.e. small stocks, growth stocks, young stocks, etc) are low. Similarly, Hvidkjaer (2008) and Barber, Odean, and Zhu (2009) find that when sentiment is high for a specific stock, subsequent returns for that stock are low. We extend this literature by documenting that when sentiment for an industry is high, subsequent returns for that industry are low. Moreover, our results suggest that firm-specific sentiment is driven largely by industry-wide sentiment.

The remainder of this paper is organized as follows. Section 2 discusses the data and presents descriptive statistics. Section 3 examines whether investors herd at the industry level. Section 4 investigates the relationship between industry proportion bought and subsequent industry returns. Section 5 decomposes the poor performance of retail investors into an industry-component and firm-specific component. Section 6 concludes.

2. Data

The data for this study come from several sources. We obtain data on returns, market capitalization, and industry classifications (SIC codes) from the Center for Research and Security Prices (CRSP). We obtain book value of equity form Compustat. We include all ordinary shares (CRSP share code 10 or 11) with adequate data. We assign each stock to one of 49 Fama and French (1997) industries.[30] Lastly, we obtain transaction data from the Institute for the Studies of Securities Market (ISSM) and the Trade and Quote database (TAQ). The ISSM dataset includes all transactions made on the NYSE and AMEX from 1983-1992 and covers NASDAQ stocks from 1987-1992. TAQ data includes all transactions from 1993 to present.

The data do not specify whether the executed trade was a buy or sell. We use the Lee and Ready (1991) algorithm to classify trades as either buyer or seller initiated. Specifically, if a trade is executed above (below) the quoted midpoint, the trade is classified as a buy (sell). If the trade is executed at the quoted midpoint, the executed trade price is compared to the preceding trade; the trade is considered a buy (sell) if the executed price was above (below) the last executed trade price. Thus all trades are classified as either a buy or a sell.

The data do not distinguish between trades made by retail investors and institutional investors. Instead, we use trade size as a proxy for individual and institutional trading. Following Barber, Odean, and Zhu (2009), trades less than $5,000 (small trades) are used to proxy for retail investor trading. Trades greater than $50,000 (large trades) are used to proxy for

[30] We use the updated industry definitions available on Ken French's website.

institutional investors.[31] Barber, Odean, and Zhu (2009) provide evidence that small trade order imbalance is positively correlated with order imbalance of retail investors at a large discount broker and a large retail full-service broker. Moreover, large trade order imbalance is negatively correlated with order imbalance from both the large discount and large retail broker, suggesting that trade size is a reasonable proxy for investor type. However, Hvidkjaer (2008) finds that many of the patterns associated with small trades disappear after 2000, presumably because it became more common for institutions to break up large orders into smaller trades after the introduction of decimalization in 2001. Consequently, this paper limits its analysis to data from 1983-2000.

In each week (month or year), from January 1983 to December 2000, for each industry, we calculate the industry proportion bought amongst retail and institutional investors. We define industry proportion bought as the number of buyer initiated trades in a given industry divided by the number of total trades in that industry. Results are very similar if we value weight each trade by the dollar volume traded.

Table 8 provides the time-series mean of cross-sectional monthly descriptive statistics. Panel A presents industry statistics. The average industry includes 98 firms, with the minimum industry containing only 5 firms and the maximum industry containing over 500 firms. The largest industry represents, on average, 10.78% of the market portfolio, while the smallest industry account for 0.08% of the market portfolio. The largest stock in an industry typically accounts for a substantial percentage of the industry's total valuation. Specifically, the largest firm accounts for roughly 30% of the average industry's market capitalization.

Panel B provides descriptive statistics on retail investor and institutional investor trading across industries. In the average industry, retail investors execute over 58,000 trades, although this ranges from 321,243 trades in the most heavily traded industries to 3,278 in the least heavily traded industries. Institutional investors execute roughly 48,000 trades in the average industry.

[31] Hereafter, we will use the term "small trader" and "retail investor" synonymously. Similarly, we will use the term "large trader" and "institutional investor" interchangeably.

Industry proportion bought exhibits substantial cross-sectional variation. Retail investors are net buyers 65% of the time in their most favored industries and only 37% of the time in their least favored industries. Similarly, institutional investor industry proportion bought ranges from 60% to 43%. The fact that retail investor industry proportion bought has a greater cross-sectional standard deviation than institutional investor industry proportion bought is consistent with our conjecture that industry herding is likely to be stronger amongst retail investors.

One concern is that retail and institutional investor trading are simple complements. Since all non-institutional investors are retail investors, and since every trade is both a buy and a sell, it seems to follow that if retail investors are herding into an industry, institutions must be herding out of the same industry. To examine this, we calculate the correlation between retail investor and institutional investor industry proportion bought. We find that the time-series average of monthly cross-sectional correlations is -0.03. This indicates that small and large trade industry order imbalances are not simple complements.

There are at least two explanations for the relatively low negative correlation between small and large trade industry proportion bought. First, our measure of small and large trade proportion bought only considers active trading through market orders. Thus passive traders who provide liquidity, either as market makers or though limit orders, are not included. This distinction is important, because a sizeable fraction of retail investor trading is done through limit orders.[32] We believe that active trades are a better measure of investor sentiment than limit orders, because whether a limit order is executed depends on the actions of others. For example, suppose retail investors have no strong belief about the technology sector and submit an equal amount of buy and sell limit orders. If institutional investors become bullish on the technology sector, then the sell limit orders of retail investors will be executed, while the buy limit orders

[32] Linnainmaa (2010), using discount brokerage data from October 2004 to September 2005, finds that limit orders account for roughly 70% of all orders placed by retail investors.

will not. In this case, the heavy sell order imbalance of retail investors simply reflects the preferences of institutional investors.[33]

The second reason our results are not complementary is because our small trading measure is meant to capture the trading of small retail investors, rather than all non-institutional investors. For example, our small trading measure is probably not very representative of the trades of very wealthy individuals. These individuals make up a sizeable portion of non-institutional trading. Wolff (2004) reports that the wealthiest 1% of households are responsible for over one-third of all US household ownership in stocks. Moreover, recent empirical evidence suggests that the trading behavior of these wealthy individuals is motivated by different considerations than the small retail traders who are the focus of this study. For example, Koirnoitis and Kumar (2010) finds that the trading behavior of retail investors with high cognitive ability (which they find is highly correlated with wealth) tends to be more motivated by information reasons, while the trading behavior of retail investors with low cognitive ability is more motivated by psychological biases.

3. Tests for industry herding

3.1 Do Investors Herd Across Industries?

In this section we examine whether the industry-wide trading of retail investors and institutional investors is systematically correlated. We first examine contemporaneous correlations. Each month we compute the proportion bought in each industry. We then calculate the Lakonishok, Shleifer, and Vishny (1992) herding measure. Let pb_{it} be equal to the proportion bought in industry i in month t and let $E[pb_{it}]$ be the expected proportion bought in month t. The herding measure for industry i in month t is computed as follows:

$$H_{it} = |pb_{it} - E[pb_{it}]| - E\big|[pb_{it} - E[pb_{it}]]\big|$$

[33] Consistent with this reasoning, Linnainmaa (2010) finds that the use of limit orders significantly alters inferences about individuals trading intentions and investment abilities.

The first term measures the difference between the proportion bought in industry i and the average proportion bought across all industries. Since the difference is an absolute value, the first term will always be non-negative. The second term in this equation is the expected value of this herding measure under the null hypothesis of no herding.[34] In essence, this equation, examines whether the realized industry proportion bought is "fat-tailed" relative to the expected industry proportion bought under the null of no industry herding.

Each month we calculate this industry herding measure for both retail and institutional investors. We average the herding measure across all 49 industries and then we take the time-series average. We find that the average industry herding measure amongst retail investors is 4.01%, while the average industry herding amongst institutional investors is 2.09%. Both measures are significantly greater than zero (p-value < .001). To get a sense of the economic importance of this effect, the 4.01% herding measure implies that if the average proportion bought was 50%, then in the average industry, 54.01% of retail trades would be on one side of the market (e.g. buying), while the remaining 45.99% of retail trades volume would be on the other side of the market (e.g. selling).

An alternative measure of herding, proposed by Sias (2004), is to examine the cross-sectional correlation between the proportion bought in period t and period $t+1$.This measure allows us to examine the persistence of investor's industry-wide preferences. Specifically, we examine the cross-sectional correlation between retail investor (institutional) industry proportion bought in week t and retail investor (institutional) industry proportion bought in week $t + x$, where x ranges from 1 week to 104 weeks. Figure 2 reports the time-series average of the cross-sectional correlations across all time periods. The correlation between retail investor industry demand this week and the prior week is over 60%. This correlation gradually declines to roughly 45% after four weeks, 34% after 12 weeks, 16% after 52 weeks, and 8% after 104 weeks. All

[34] Since pb_{it} follows a binomial distribution, the expected value of this measure can be computed for any given average proportion bought (i.e. the probability of success) and the number of trades.

estimates are significantly greater than zero.[35] Thus retail investor industry trading is not only contemporaneously correlated but also highly persistent. Moreover, across all horizons, the cross-sectional correlation between retail industry herding is typically 2 to 3 times as large as the cross-sectional correlation of institutional industry herding.

3.2 Is Industry Herding Driven by Stock Herding or Size and Book-to-Market Herding?

Table 8 indicates that, on average, the largest firm in an industry accounts for roughly 30% of the industry's market capitalization and roughly 27.5% of retail investors total trades. Thus, one concern is that our industry-level results are being driven by investors herding into the largest stock in the industry. An additional concern is that industry herding may be due to the fact that stocks in the same industries tend to have similar characteristics such as size and book-to-market ratios. For example, technology stocks tend to be growth oriented, while utility stocks tend to be value stocks. Teo and Woo (2004) and Kumar (2009) provide evidence that investors tend to categorize stocks based on size and book-to-market. Thus, it is worth examining whether industry herding persists after controlling for firm-level herding and herding into stocks with similar size and book-to-market ratios.

To examine this issue, we run Fama-Macbeth regressions where the dependent variable is the proportion bought in stock *i* in week t. We then include three independent variables. The first is the proportion bought in stock *i* in week *t-x*. This variable captures industry herding that is due to firm-level herding. We then assign all other stocks in the same industry to one of six size and book-to-market styles based on the Fama and French (1993) methodology. The second variable is the average industry proportion bought amongst stocks in the same industry and size and book-to-market styles in week *t-x*. The third variable is the average industry proportion bought amongst stocks in the same industry but in a different size and book-to-market style in week *t-x*. Thus, this

[35] Standard errors are computed from the time-series average. We find that the estimates are significantly serially correlated for the first few lags but that this correlation declines quickly. For example, at one lag, the serial correlation is 0.36, but this declines to 0.09 at lag 6. To adjust for this serial correlation, we use Newey-West standard errors with six lags. Using additional lags does not significantly alter the standard errors.

last variable reflects industry herding that is distinct from firm-level herding and size and book-to-market style herding.

Panel A of Table 9 reports the time-series average of the cross-sectional estimates for retail investors. The t-statistics are computed using the Newey-West (1987) correction.[36] The first column reports results based on the proportion bought over the prior week. Retail investor demand for a stock this week is strongly related to their demand for the stock last week. A 10% increase in the proportion bought of a stock in the prior week will increase the expected proportion bought of the stock by 2.8% in the following week. Retail investor demand for a stock this week is also related to their demand for stocks with similar size and book-to-market styles in the same industry. However, even after controlling for these effects, we see that retail investor demand for a stock is strongly related to its demand for other stocks in the same industry with different size and book-to-market styles. The second, third, and fourth column indicates that retail investor demand for a stock is positively related to their industry-wide demand measures over the prior 2 months, 6 months, or a year. The results indicate that industry herding is distinct from firm-level herding and size and book-to-market style herding and provide further evidence that retail investor industry herding is highly persistent.

Panel B of Table 9 repeats the analysis for institutional investors. Like retail investors, institutional investors demand for a stock is positively related to their demand for the stock in the prior week, along with their demand for other stocks in the same industry in the prior week. However, the relationship is weaker amongst institutional investors. The adjusted R^2 from the institutional regressions is roughly half the adjusted R^2 from the retail investor regressions. Moreover, the coefficients for institutional investors are always less than half the magnitude of those for retail investors. Institutional investor industry herding is also less persistent. There is no

[36] Unless otherwise specified, Newey-West standard errors are computed using six lags. Using more than six lags does not significantly alter the standard errors in Table 2.

significant relationship between institutional demand for a stock and its demands for other stocks with in the same industry over the past 6 months to 1 year.

3.3 Prior Returns and Industry Proportion Bought

The previous results establish that retail investors have strong and persistent preferences for certain industries. The style investing model of Barberis and Shleifer (2003) posits that these preferences may be related to prior returns. Specifically, Barberis and Shliefer (2003) model an economy in which there are fundamental traders and "switchers". These switchers move their wealth out of poorly performing styles and into styles that have performed well. This implies that industry proportion bought will be positively related to past industry returns. To examine this implication, each month, from January 1983 to December 2000, we run the following cross-sectional regression:

$$IND\_PB_{it} = a_o + b_1 IndSize_{it} + b_2 IndBM_{it} + b_3 IndRet_{it-1} + b_4 IndRet_{it-3,t-2} + b_5 IndRet_{it-6,t-3} + b_6 IndRet_{it-12,t-7} + b_7 IndRet_{it-24,t-12} + b_8 Ind\_PB_{it-1} + b_9 Ind\_PB_{it-3,t-2} + b_{10} Ind\_PB_{it-6,t-3} + b_{11} Ind\_PB_{it-12,t-7} + b_{12} Ind\_PB_{it-24,t-12}$$

The dependent variable is the industry proportion bought. The independent variables include IndSize and IndBM which are equal to the industry average size and the industry average book-to-market ratio (both in natural logs). We then include several measures of prior industry returns, ranging from the prior one month return to the return over the prior 12 to 24 months. In addition, we include lagged levels of industry proportion bought.

Panel A of Table 10 reports the time-series average of the cross-sectional results. Standard errors are computed using the Newey-West correction. The first column of panel A reports the results for retail investors. Industry proportion bought is negatively related to industry returns over the prior 3 months. This suggests that retail investors do not immediately withdraw assets from poorly performing styles and invest in recent winning styles. However, industry proportion bought is positively related to prior industry returns over the past 4 to 6 months, 7 to

12 months, and 13 to 24 months. The impact of prior industry returns on industry proportion bought is both statistically and economically significant. For example, a 10% increase in the industry return over the prior 13 to 24 months would increase the industry proportion bought by 6.5%. We also examine whether prior industry returns can forecast industry proportion bought, after controlling for lagged industry proportion bought. The results of column 3 indicate that both prior industry return and prior industry proportion bought are significantly related to industry proportion bought.

Columns 5 and 7 repeat the analysis for institutional investors. Unlike retail investors, institutional investors are significant short-term industry momentum traders. The industry return over the prior month positively forecasts institutional industry proportion bought. This result persists even after controlling for institutional industry proportion bought over the prior month. However, there is no significant relationship between institutional industry proportion bought and industry returns over the prior 2 to 12 months.

A question of interest is whether style-level momentum trading is distinct from firm-level momentum trading. To address this question, we examine whether prior industry returns can forecast firm-level proportion bought after controlling for firm-level prior returns. Thus, the dependent variable of this regression is the firm-level proportion bought and all the independent variables are firm-level variables with the exception of industry returns. Panel B of Table 22 reports the time-series average of the monthly coefficients for this regression. Consistent with Hvidkjaer (2006), we find that retail investors tend to be firm-level contrarians over short horizons, but firm-level momentum traders over longer horizons. Moreover, after controlling for firm-level returns, industry returns now positively forecast firm-level proportion bought across all horizons. Thus over shorter horizons both firm-level and industry-level returns can forecast firm-level proportion bought but in opposite directions. The results suggest that prior industry performance and prior firm-level performance influences the investment decisions of retail investors in a fundamentally different way.

The fifth column of Panel B analyzes the firm-level and industry-level momentum trading of institutional investors. Consistent with many prior studies on institutional investors, our result indicate that institutional investors are firm-level momentum traders.[37] In addition, firm level proportion bought is significantly positively related to prior one month industry returns. However, firm-level proportion bought is significantly negatively related to industry returns over the prior 6 to 24 months. Thus, unlike retail investors, institutional investors are not industry-level momentum traders.

4. Industry Herding and Industry Returns

The results suggest that retail investors herd into winning industries and herd out of losing industries. The style investing model posits that this herding is motivated, at least in part, by investor sentiment. Moreover, it argues that this sentiment related demand cannot be completely offset by the actions of rational arbitragers. Consequently, the style investing model predicts that style-level sentiment pushes prices away from fundamentals in the short run, leading to long-term reversals. These predictions are in sharp contrast to the rational herding models of Froot, Sharfstein, and Stein (1992) and Hirshleifer, Subrahmanyam and Titman (1994) which argue that herding is driven by investors receiving correlated signals about fundamental information. These rational herding models argue that herding pushes price towards fundamentals, and therefore, do not predict subsequent price reversals.

4.1 Industry Herding and Industry Returns

To explore these competing explanations, we sort industries into quintiles based on retail investor (or institutional investor) industry proportion bought and examine their subsequent returns. The first trading strategy we consider is to sort on prior 3 month industry proportion bought and hold that portfolio for 3 months (3m-3m strategy). For example, from April 1983 to June 1983, portfolio 1 (5) would consist of the quintile of industries most heavily sold (bought)

[37] Studies that provide evidence of institutional momentum trading include Lakonishok, Shleifer and Vishny (1992), Grinblatt, Titman, and Wermers (1995) and Badrinath and Wahal (2002).

by retail investors from January 1983 to March 1983. For each portfolio, we first compute the value-weighted performance of each industry in the portfolio.[38] We then take the equally weighted average of each industry's return in that portfolio.[39] This gives us a time series of monthly returns starting in April of 1983 and ending in December of 2000.

Panel A of Table 11 reports the average monthly market-adjusted returns for each quintile. Market-adjusted returns are the difference between the portfolio return and the value-weighted market index. Interestingly, the industries most heavily bought by retail investors in the prior quarter underperform the industries most heavily sold by retail investors over the subsequent 3 months by roughly 48 basis points (bps) per month. This is estimate is highly statistically significant and translates into an annual outperformance of nearly 6%. In contrast, the industries most heavily bought by institutional investors outperform the industries most heavily sold by about 16 bps per month; however this estimate is not significantly different from zero.

To see if the poor performance of retail investors is driven by retail investors loading on factors with poor performance, we also compute five-factor alphas for each portfolio. We compute five-factor alphas using a time-series regression. The dependent variable is the monthly return on a given portfolio less the risk-free rate, and the independent variables represent factors related to market, firm size, book-to-market, firm-level momentum, and industry momentum. The first four factors are taken from Ken French's data library.[40] The fifth factor is included to control for the industry momentum effect documented by Moskowitz and Grinblatt (1999).[41] The five-factor alpha results indicate that a portfolio that went long the industries most heavily bought by retail investors and short the industries most heavily sold by retail investors, would earn a

---

[38] Equally weighting each stock in the industry yields stronger results.

[39] We equal weight each industry. Value weighting each industry leads to very similar conclusions.

[40] See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html for more details on the construction of these factors.

[41] To construct the industry momentum factor, we use six value weighed portfolios formed on average industry size and prior 12 month industry returns. The portfolios, which are formed monthly, are the intersection of 2 portfolios formed on size and 3 portfolios formed on prior industry returns. Industry momentum is the average return on the two high prior return portfolios minus the average return of the two low prior return portfolios.

monthly alpha of -41 bps. This estimate remains highly significant and indicates that factor loadings cannot explain the poor industry selection of retail investors. The five-factor alpha for the long-short portfolio based on institutional industry proportion bought is 11 basis points, and is not statistically significant.

We explore several other strategies. Panel B of Table 11 presents results for a trading strategy that sorts on prior 6 month industry proportion bought and then holds the portfolio for 6 months (6m-6m). Similarly, Panel C shows the results of a trading strategy that sorts on prior 12 month industry proportion bought and then holds the portfolio for 12 months (12m-12m). The results from these strategies are very similar to the 3m-3m strategy. In both cases, the industries most heavily bought by retail investors significantly underperform the industries most heavily sold by retail investors. The long-short portfolio for the 6m-6m strategy earns a five-factor alpha of roughly -39 bps per month, and the long-short portfolio for the 12m-12m strategy earns a five-factor alpha of roughly -34 bps per month. In unreported results, we also consider strategies shorter than 3 months and longer than 12 months. We find that a 1m-1m strategy earns a five-factor alpha of roughly -33 bps which is marginally significant (t-stat = -1.89) and that a 24m-24m strategy earns a five-factor alpha of -10 bps which is not significantly different from zero. For all holding periods and formation periods, the long short portfolio based on institutional industry proportion bought does not earn returns that are significantly different from zero. These results seem most consistent with retail investor industry demand being driven, at least in part, by sentiment; while institutional industry demand is more driven by fundamentals.

4.2 Industry Herding and Weekly Returns

The style investing model of Barberis and Shleifer (2003) not only predicts long-term reversals but also short-term style-level momentum. This section investigates whether there is any evidence of style-level momentum at shorter horizons. First, however, we examine the relationship between industry proportion bought and contemporaneous returns. Each week, from 1983-2000, we sort industries into quintiles based on small and large trade industry proportion bought. For each industry, we compute the value-weighted return for each day during the sorting week. We then calculate the performance of each quintile by taking the average of each industry's return within the quintile. We compound the daily returns to obtain a monthly return series. Panel A of Table 12 presents the results of this analysis. Industry returns are strongly related to both small and large trade proportion bought, although the magnitude is significantly larger for institutional proportion bought. This is consistent with the larger trades of institutional investors having significantly greater price impact than the smaller trades of retail investors. However, this is also consistent with institutional investors being significantly greater short-term industry momentum traders. We do not attempt to determine the causality of this relationship.[42]

Next we examine whether industry proportion bought can forecast the subsequent week's industry returns. Each week we sort industries into quintiles based on the retail investors (or institutional investor) proportion bought. The value-weighted return for each industry is computed over the subsequent five trading days. Each day, we calculate the performance of each quintile by taking the average of each industry's return in that quintile. Thus, we obtain a time series of daily returns. We compound these daily returns into monthly returns.

Panel B of Table 12 reports the market-adjusted and five-factor alphas for the portfolios sorted on prior week industry proportion bought. Consistent with Barberis and Shleifer (2003), amongst retail investors, we find strong evidence of industry-level continuations. A portfolio that went long the industries most heavily bought by retail investors in the prior week and short the industries most heavily sold by retail investors would earn a five-factor alpha of 62 basis points a month. This effect is highly statistically and economically significant and runs counter to the typical pattern of short-term reversals documented by Jegadeesh (1990) and Lehman (1990). These short-term continuations (in conjunction with long-term reversals) are consistent with

[42] Prior research that investigated the relationship between order imbalance and contemporaneous returns has found evidence that supports both explanations (see e.g Griffin, Harris, and Topalogu (2003) and Sias, Starks, and Titman (2006)).

persistent retail investor industry-wide sentiment pushing prices away from fundamental values in the short run.

Table 24 also reports the results for institutional industry herding. In contrast to our retail investor results, here we find strong evidence of industry reversals. Specifically, a portfolio that went long the industries most heavily bought by institutional investors over the prior week and short the industries most heavily sold would earn a five-factor alpha of -45 basis points. The magnitude of this reversal is relatively small compared to the contemporaneous price effects associated with institutional industry proportion bought (roughly 858 bps). This result is consistent with large institutional traders requiring short-term liquidity. This price pressure temporarily pushes prices up leading to short-term reversals. An alternative explanation is that institutional investors overreact.

4.3 Fama-Macbeth Weekly Regressions

As an additional test, we examine how weekly industry returns are a function of industry proportion bought over the prior two years. For both retail and institutional investors, each week, we estimate the following cross-sectional regression:

$$\begin{aligned} IndRet_{it} = a_o &+ b_1 Ind\_PB_{it-1} + b_2 Ind\_PB_{i,t-4,t-2} \\ &+ b_3 Ind\_PB_{i,t-8,t-5} \sum_{w=9}^{97\ by\ 8} b_{t-w,t-w-7}\, Ind\_PB_{it-w,t-w-7} + c_1 MVE_{it} \\ &+ d_1 BM_{it} + \sum_{w=1}^{4} e_{t-w,}\, Ind\_Ret_{it-w,} + f_1 Ind\_Ret_{it-52,t-5} \\ &+ g_1 Ind\_Ret_{it-104,t-53} + \varepsilon_i \end{aligned}$$

The dependent variable is the industry return in week *t.* The independent variables include the industry proportion bought over the prior week, prior two to four weeks, prior 5 to 8 weeks, and subsequent 8 week periods, beginning with the prior 9 to 16 weeks and ending over the prior 97 to 104 weeks. We also include controls for factors that are known to influence industry returns. We include the average industry market cap (the natural log of the market value of equity) and the average industry book-to-market ratio (in natural logs). Lastly, to control for the industry momentum, we include variables to capture past industry returns over different horizons.

Figure 3A plots the coefficient estimates for lagged industry proportion bought by retail investors. The coefficients are based on the time-series average of the cross-sectional estimates. Standard errors are based on the time-series standard deviation of the weekly estimates. The stanrdard errors are adjusted for serial correlation using the Newey-West (1987) correction. The figure indicates that the industry proportion bought by retail investors over the prior 4 weeks positively forecasts returns. In contrast, industry proportion bought over the past 9 week to the past 72 weeks consistently negatively forecasts returns. The results provide additional support for the style investing model of Barberis and Shleifer (2003).

Figure 3B graphs the results for prior industry proportion bought by institutional investors. Consistent with our weekly results, industry proportion bought over the prior week is negatively related to industry returns. However, over longer horizons there is no consistent relationship between prior industry proportion bought and industry returns.

5. Industry Sentiment vs. Firm Sentiment

Barber, Odean, and Zhu (2009) and Hvidkjaer (2008) document that the stocks bought by retail investors underperform the stocks sold by retail investors. This section investigates to what extent the poor firm-level performance of retail investors is driven by their poor industry selection. To examine the issue, we repeat the industry analysis of Tables 23 and 24, but substitute stock proportion bought for industry proportion bought. In other words, for each strategy we sort stocks based on retail investor firm-level proportion bought over the past n months (where n can equal 1 week, 3 months, 6 months, or 12 months) and then hold that portfolio for n months. The return on that portfolio is the value-weighted return of each stock in that portfolio. We then decompose the performance of this portfolio into industry performance and firm-level performance. Following Busse and Tong (2009), we compute industry performance by replacing each stock in the quintile with its value-weighted industry return. The industry return receives the same weight as the stock it represents in the portfolio. This measure is a proxy for the performance of retail investors that is due to their industry selection. The difference between their total performance and this industry performance is a measure of retail investor's performance due to their stock selection.

For example, suppose Microsoft made up 80% of quintile 1 and Goldman Sachs made up the remaining 20% of quintile 1. Suppose Microsoft earned 3%, Goldman Sachs earned 2%, the tech industry earned 1%, and the financial industry earned 4%. Under this scenario, quintile 1's total performance would be 2.8%, its industry return would be 1.6% and its firm return would be 1.2%.

Table 13 reports the results of this decomposition. Panel A reports the results for the 1w-1w strategy. Consistent with Barber, Odean, and Zhu (2009) the total performance of retail investors is significantly positive over this horizon. A portfolio that went long the stocks most heavily bought by retail investors and short the stocks most heavily sold by retail investors would earn an average monthly five-factor alpha of 79 bps. The decomposition indicates that the industry selection is responsible for roughly 43 bps (54%), while the stock selection is responsible for 37 bps (46%). Both the industry component and stock level component contribute significantly to the short-term momentum.

Consistent with both Hvidkjaer (2008) and Barber, Odean, and Zhu (2009), Panels B, C, and D all document a negative relationship between retail investor firm-level proportion bought over the prior quarter, six months, or a year, and subsequent firm-level returns. For example, Panel C indicates that a portfolio that went long the stocks most heavily bought by retail investors and short the stocks most heavily sold by retail investors over the prior six months, would earn an average monthly five-factor alpha of -54 bps over the subsequent six months. The decomposition indicates that roughly 63% (34 bps) of total underperformance is due to retail investors' industry-wide selection, while 37% (20) bps is due to their firm-level selection. Moreover, the industry component remains reliably different from zero indicating that the industry selection of retail investors contributes significantly to their overall poor performance. In contrast, the firm-level component is no longer significantly different from zero. The 3m-3m and 12m-12m decomposition results yield similar conclusions.

6. Conclusion

This paper examines the industry-wide investment decisions of retail and institutional investors. We find that the industry-wide trading behavior of retail investors is consistent with the style investing model of Barberis and Shleifer (2003). Specifically, we find that retail investors herd into and out of the same industries and that their herding is highly persistent. Retail investors prefer industries with high returns over the past two year. Retail investor herding has a strong impact on contemporaneous prices and also positively forecasts returns over the subsequent week. Over longer horizons, however, retail investors' industry proportion bought negatively forecasts industry returns. Thus, retail investors appear to behave very much like the "style switchers" described in Barberis and Shleifer (2003). They chase industries that have done well in the past, pushing prices away from fundamentals.

Our finding that retail investor industry proportion bought forecasts industry returns are similar to the findings of Barber, Odean, and Zhu (2009) and Hvidkjaer (2008) who document that retail investor firm-level proportion bought forecasts firm returns. To assess the extent to which the poor performance of retail investor trading is driven by their industry-wide investment decisions, we decompose the performance of retail traders into an industry component and a firm-specific component. Our industry decomposition reveals that roughly 60% of the poor performance is driven by the poor industry selection of retail investors. Moreover, this industry component remains significantly negative, while the firm-specific component is no longer reliably different from zero. Taken together, our findings suggest that retail investors categorize stocks by industry and that industry-wide sentiment accounts for a substantial portion of the poor performance of retail investors.

We also find that institutional investors exhibit industry herding. However, the magnitude and persistence of institutional industry herding is smaller than that of retail investor industry herding. Moreover, institutional industry herding does not seem to be well described by the style investing model. Institutional industry herding is associated with very large contemporaneous price increases, a small portion of which reverses in the subsequent week. Over longer horizons, institutional industry proportion bought is not significantly related to industry returns. The fact that institutional industry herding does not generate long run reversals, suggests that unlike retail investors, institutional investor herding is not motivated by sentiment. Institutional industry herding seems better described by rational herding models such as Froot, Sharfstein, and Stein (1992) and Hirshleifer, Subrahmanyam and Titman (1994).

Third Essay: Mutual Fund Industry Selection and Persistence

1. Introduction

Mutual fund studies typically analyze fund performance either at the fund level or at the individual security level. At the fund level, shareholder returns are usually compared to one or more benchmarks, such as the S&P 500. At the security level, individual stock returns are evaluated relative to stock-specific benchmarks. Examples of the former range from the earliest mutual fund studies, including Jensen (1968), up to the present. Examples of the latter include Grinblatt and Titman (1989) and Wermers (2000), among many others.

The interpretation of these performance studies invariably emphasizes the fund manager's stock-picking ability. For instance, a positive alpha suggests that the manager has stock-picking skill. However, the specific reason why a fund manager holds a top-performing stock can go far beyond his ability to pick individual stocks. For example, a manager may have skill to interpret the economy and shift his portfolio towards the types of stocks that do well during certain macroeconomic environments. When interest rates begin to decrease, for instance, banks tend to outperform as their margins improve.

The stock-picker label seems most appropriate for those that employ a bottom-up investment technique. In this type of approach, the manager focuses on the analysis of individual companies and de-emphasizes economic cycles and industry trends. The alternative to the bottom-up investment style is the top-down approach. In this approach, managers first make decisions regarding broad industry allocations before moving on to the finer details and eventually selecting individual stocks.

In this paper, we explore manager skill in making decisions regarding broader allocations. Specifically, we examine the relative importance of industry selection compared to stock selection in the performance of a manager's portfolio. That is, we examine the extent to which a manager's industry allocations drive his performance vs. his specific stock choices within the industries held in his portfolio. Top-performing managers may do well because they choose stocks in top-performing industries, where average stocks in those same top industries would have performed just as well as the stocks chosen by the managers. Alternatively, top-performing managers may choose the best stocks in average or even underperforming industries.

We show that industry selection contributes substantially to fund performance, accounting for roughly one fourth of a fund's abnormal performance based on two-digit Standard Industry Classification (SIC) codes.

The skill sets associated with industry- and stock-selection ability could differ considerably, with industry-selection ability relying on understanding macroeconomic relationships, and individual-stock-selection skill relying on the ability to size up firm-specific drivers, such as innovative products or managerial competence. We analyze the extent to which each component of skill persists. Numerous papers examine the extent to which overall skill persists, including Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Goetzmann and Ibbotson (1994), Brown and Goetzmann (1995), Malkiel (1995), Elton, Gruber, and Blake (1996), Carhart (1997), and Bollen and Busse (2005). We find that the industry-selection component of performance shows stronger persistence over longer time horizons than the stock-selection component of performance. Whereas past industry selectivity predicts future industry selectivity for investment horizons up to three years, stock selectivity does not persist beyond one year. Our results suggest that industry selection, rather than stock selection, drives the evidence of overall performance persistence documented in the literature.

Berk and Green (2004) hypothesize that the large flows of capital into successful funds eventually lead to the successful funds losing their performance edge. Successful funds face increasing transaction costs (due to greater-size trades) and/or the addition of less attractive stocks to their portfolio. Consistent with the results of Chen et al. (2004) for overall performance, we find a negative relation between fund portfolio size and stock-selection skill. By contrast, we find no evidence of a negative relation between fund size and industry-selection skill. Although funds are unable to maintain their stock selectivity when their assets increase, they do maintain their industry-selection ability at large levels of assets. Thus, flows into successful funds do not appear to erode industry skill. Apparently, unlike individual stocks, industries provide ample opportunities for further investments.

Examining the industry features of fund portfolios has received little attention among mutual fund studies. Rather than controlling for industry exposure, most mutual fund studies control for exposure to size, value, and momentum factors in their performance measures, consistent with trends in the empirical asset pricing literature. Recent papers that examine industry allocations include Kacperczyk, Sialm, and Zheng (2005) and Avramov and Wermers (2006). Kacperczyk, Sialm, and Zheng (2005) find that funds that concentrate their holdings in fewer industries tend to outperform funds that diversify more across industries. Avramov and Wermers (2006) examine the industry allocations of funds predicted to outperform based on manager skill, risk loadings, and benchmark returns. They find that optimally-chosen funds show ability to time industry allocations across the business cycle and have larger exposure to the energy, utilities, and metals industries.

In a paper widely referenced by practitioners, Brinson, Hood, and Beebower (1986) explore the importance of allocations one step higher in the investment process for portfolios managed by institutional money managers. They analyze allocations among stocks, bonds, and cash, and find that these allocation decisions explain more than 90 percent of the variation in a portfolio's total return. By construction, our sample of mutual funds already primarily holds equities. Consequently, we begin at the industry, rather than asset-class, level. Furthermore, we focus on determining the extent to which industry allocations explain risk-adjusted performance, rather than variation in total return.

The paper proceeds as follows. Section 2 describes the data. Section 3 defines our measures of industry and stock selection. Section 4 presents our empirical analysis, including performance persistence and issues related to scale. Section 5 concludes the paper.

2. Data

We obtain mutual fund holdings from Thomson Financial's CDA/Spectrum Mutual Fund Holdings database. The database consists of quarterly stockholdings data for virtually all U.S. mutual funds between January 1980 and December 2006 (inclusive), with no minimum survival requirement for a fund to be included. For each stock holding of each fund, the data include CUSIP, ticker symbol, company name, and number of shares held. Thomson Financial collects these data both from reports filed by mutual funds with the SEC, as required by amendments to Section 30 of the Investment Company Act of 1940, and from voluntary reports generated by the funds. Although mutual funds have been required to file holdings reports with the SEC on a semi-annual basis since 1985, quarterly reports were obtained from more than 80 percent of funds during most of the 1985 to 2006 time period. Prior to 1985, more than 90 percent of funds reported on a quarterly basis.

We focus on domestic equity funds and include those with the following investment objective codes as indicated by Thomson Financial: Aggressive Growth, Growth, and Growth & Income. Since we are interested in analyzing the skill associated with actively managed funds, we remove funds that are likely to be passively managed.[43] We also remove sector funds because their industry allocation decisions are substantially constrained.[44]

We obtain individual stock returns, prices, shares outstanding, and SIC codes from the Center in the Research of Security Prices (CRSP) Daily and Monthly Stock files. We collect the data from CRSP for the 27-year sample period from 1980 to 2006.

We match the stock holdings from Thomson Financial with the daily stock returns from CRSP. Although we are unable to match all stock holdings to companies listed in CRSP, the missing data constitute less than one percent of the stock holdings, which is consistent with the match rate of Kacperczyk, Sialm, and Zheng (2005). Since CRSP

---

[43] The standard research databases do not identify passively managed funds. Consequently, our approach for removing these funds is imperfect. We remove from the sample funds whose names contain any of the following text strings: *Index*, *Ind*, *Idx*, *Indx*, *Mkt*, *Market*, *Composite*, *S&P*, *SP*, *Russell*, *Nasdaq*, *DJ*, *Dow*, *Jones*, *Wilshire*, *NYSE*, *ishares*, *SPDR*, *HOLDRs*, *ETF*, *StreetTRACKS*, *100*, *400*, *500*, *600*, *1000*, *1500*, *2000*, *3000*, *5000*. This procedure eliminates roughly five percent of the sample.

[44] The standard research databases also do not identify sector funds. We remove funds whose names contain text strings typically associated with a sector fund, such as *bank*, *mining*, and *tech*, among many others. This procedure removes approximately six percent of the sample.

focuses on U.S. listed stocks, which are required to meet minimum market capitalization requirements, unmatched holdings likely consist mainly of micro-cap and foreign stocks not listed on the NYSE, Amex, or NASDAQ. In untabulated results, we find that our main findings are not sensitive to the degree of match.

We compute gross daily fund returns using portfolio weights derived from the most recent portfolio holdings snapshot. For example, we compute returns for 1990Q3 using portfolio holdings from the end of June 1990. We determine each portfolio holding's initial weight by taking the product of stock price and shares held and then dividing this dollar investment by the sum of the dollar investments across all stocks in the portfolio. The weights evolve during the quarter as they would in a buy-and-hold portfolio, where weights change daily as a function of the returns of all portfolio holdings. When additional holdings data become available during a calendar quarter, we reset the individual holding weights beginning the day after the date of the new holdings data using the new shareholdings and stock prices. Otherwise, we assume shareholdings are constant through the quarter. In instances where holdings are not reported quarterly (e.g., semi-annual holdings) or do not align with calendar quarters, we use the most recently reported shareholdings. The calendar time between a daily return estimate and the fund holdings it is derived from is at most six months (and typically less than three months).

The procedure that we use to compute returns is similar to that used by others, such as Grinblatt and Titman (1989) and Wermers (2000). The return series differ from actual shareholder returns because they ignore expenses, transaction costs, non-U.S.

equity holdings, and intra-quarter portfolio adjustments. The extent to which these differences affect our results is unclear, although no specific bias is obvious.

Table 14 provides portfolio statistics of our fund sample for select years during our sample period. The number of funds increases dramatically from 1980 through 2006, consistent with the explosive growth in the mutual fund industry over the last 25 years. The number of stocks per portfolio also increases considerably during the sample period, coinciding with an increase in average assets under management per fund. Increasing the number of stocks in a portfolio can help to mitigate the increase in transaction costs that would normally accompany an increase in assets.

The table also reports the number of industries per portfolio, where we use the two-digit SIC code to define industries, taken from CRSP.[45] SIC codes are typically used at the two- or four-digit level. We use the coarser two-digit level in our analysis for two main reasons. First, fine industry groupings (such as those associated with four-digit SIC codes) often lead to sparsely-populated industries, making it difficult to disentangle the industry effect from the individual stock effect. For example, Microsoft Corp. (ticker MSFT) accounts for 39 percent of the total market capitalization associated with the 7370 SIC code during the 1986-2006 time period. During periods of time when MSFT stock outperforms the market, the entire four-digit software industry typically does well also (due to MSFT's direct weighting in the industry's returns), and MSFT's performance beyond the industry is muted. Coarser industry groupings lead to more heavily-populated industries and smaller individual stock influences. A second and less quantifiable reason is that most diversified fund managers are unlikely to categorize their holdings into

[45] We use historically accurate SIC codes, rather than header (most recent) SIC codes.

groups much beyond the equivalent of two-digit SIC codes. Fund managers, for instance, often incorporate information from third-party data providers as one input into their analysis, and many of these data are organized into industry groupings of similar granularity to that of two-digit SIC codes.[46]

Alternatives to the SIC classification system include North American Industry Classification System (NAICS) codes and Global Industry Classification System (GICS) codes. We choose the SIC system because of its widespread use in the academic literature. However, as a robustness test, we repeat our main analyses with three-digit NAICS codes, which we take from Compustat, and find very similar results. GICS codes are not available until the mid-1990's, and would therefore be unavailable for more than half of our sample period.

A total of 95 unique two-digit SIC codes exist, ranging from 01 (Agricultural Production - Crops) to 99 (Nonclassifiable Establishments).[47] At any point in time, our sample funds in aggregate hold stocks in about 95 percent of these two-digit SIC codes. With roughly 8,000 stocks in operation and available on CRSP at any given point in time during our sample period, an average of about 80 CRSP stocks exist per specific two-digit SIC code. As indicated in Table 1, each fund in our sample holds a median of 52 stocks in a median of 22 unique two-digit SIC industries. Thus, on average, funds hold roughly three percent of the stocks within the industries included in their portfolio.[48]

---

[46] For example, Ned Davis Research, one of the most widely-subscribed to investment research services for institutional money managers, divides stocks into 115 "sub-industry" groups, while Standard & Poor's uses a total of 145 industry groups.

[47] There are 99 unique three-digit NAICS codes.

[48] 52/22 stocks held per industry out of 80 stocks total per industry.

3. Performance Decomposition

Here and elsewhere in the paper, we evaluate performance using three different standard base models: a one-factor model, based on the capital asset pricing model, that uses the excess returns on a proxy for the overall stock market as the factor; the three-factor model that uses size (SMB) and value (HML) factors together with the market factor (see Fama and French (1993)); and the four-factor model that adds a momentum (UMD) factor to the three-factor model (see Jegadeesh and Titman (1993) and Carhart (1997)):

$$r_{p,t} = \alpha_p + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t}, \tag{1}$$

where $r_{p,t}$ is the excess return of a fund portfolio at time *t*, and $r_{j,t}$ are the returns of the k = 1 to 4 factors. We use the value-weighted CRSP return series for our market proxy, and take the SMB, HML, and UMD factors and the risk free return (to compute excess portfolio returns) from Ken French's website. The intercept, α_p, is a standard estimate of mutual fund skill, and it captures the ability of funds to outperform the market on a risk-adjusted basis, with adjustments for size, value, and momentum anomalies in the three- and four-factor models.

We interpret the standard estimate of skill, alpha, as the sum of two distinct components of skill: industry-selection skill and individual-stock-selection skill. Hereafter, we use industry-selection skill synonymously with industry alpha and individual-stock-selection skill synonymously with stock alpha. Industry-selection skill is the ability to allocate assets to industries that subsequently outperform other industries.

For many fund managers, industry-selection skill captures expertise in one of the early steps in the investment process—the ability to choose the broad areas of the market that will outperform. Individual-stock-selection skill is the ability to pick the best stocks within the industries in which a fund invests.

We decompose standard alpha into industry and stock alphas as follows. First, for each fund, we construct a corresponding time series of industry returns, $R_{pi,t}$, consistent with the fund's industry exposures. To do so, we replace each stock in the fund's portfolio with its value-weighted industry return. Thus, we replace Microsoft, for example, by the value-weighted return associated with two-digit SIC industry 73, which is Microsoft's two-digit SIC industry assignment. Each industry return receives the same weight as the stock it represents in the fund portfolio. Thus, this new time series of returns strips out the dynamics of individual stocks, leaving only that which is attributable to the fund's industry exposures.

We use this fund-specific industry time series two different ways. First, we use its excess returns as a regressand in a regression similar to equation (1),

$$r_{pi,t} = \alpha_{pi} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t}, \tag{2}$$

where $r_{pi,t} = R_{pi,t} - r_{f,t}$. We interpret the intercept in these models, α_{pi}, as fund industry-selection skill, the ability to allocate assets to industries that outperform other industries. Second, we orthogonalize each fund's excess industry return series with respect to the factors in regression equation (1) and then include the orthogonalized factor, $r^{o}_{pi,t}$, as an additional regressor:

$$r_{p,t} = \alpha_{ps} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \beta_{pi} r^{o}_{pi,t} + \varepsilon_{p,t} . \tag{3}$$

We interpret the intercept, α_{ps}, as individual-stock-selection skill, the ability of funds to pick stocks that outperform other stocks in the same industries held by the fund.[49] Note that regression equation (3) is fund specific, since each fund has a different final factor tailored to its own unique industry exposures.

We estimate regression equations (1), (2), and (3) each quarter, and take the mean of the skill estimates each quarter and then across quarters. Table 15 shows the mean α_p, α_{pi}, and α_{ps} and *t*-statistics (based on the standard error of the mean of the time series of mean quarterly alphas) for the base one-, three-, and four-factor regression models.[50] The mean estimates of overall skill are positive and statistically significant. Recall that our returns are gross of expenses and transaction costs, so it is perhaps not surprising that these results are not directionally consistent with the results of studies that examine shareholder returns (net of expenses and transaction costs). Examining shareholder returns typically leads to evidence of negative risk-adjusted performance (see, for example, Gruber (1996)). Our evidence of a positive mean overall skill estimate is similar to the results of Wermers (2000), who finds evidence of positive mean gross performance net of DGTW benchmarks (Daniel et al. (1997)).

---

[49] $r^{0}_{pi,t}$ is an orthogonalized industry factor (removing the market, SMB, HML, UMD effect from $r_{pi,t}$). β_{pi} measures the sensitivity of fund's returns to the relevant industry returns. Using $r^{0}_{pi,t}$ rather than $r_{pi,t}$ does not affect the estimate of individual-stock-selection skill and allows one to interpret the industry sensitivity as an incremental effect beyond market and style (SMB, HML, and UMD) exposures.

[50] First-order and higher-order autocorrelations in the time series of mean quarterly alphas are not statistically significant.

The table also shows positive mean estimates for the industry and stock components of alpha for all three regression models. However, these are not statistically significant. Our main goal in examining these skill estimates is to provide an initial indication of the relative importance of the industry component of alpha. Based on the sample means, industry-selection skill appears to drive about a quarter of the fund's overall performance. For example, for the four-factor model, industry-selection skill is 25.5 percent of total skill (0.0012 percent out of 0.0012+0.0035 percent).[51] However, the lack of statistical significance in the estimates of industry- and individual-stock-selection skill suggests that this initial estimate should be interpreted with caution, perhaps pointing in the direction of the importance of industry selection skill, but not representing a precise estimate.

In a result not shown in the table, the mean cross-sectional correlation between the industry and stock components of alpha is 0.04.[52] The small correlation suggests that the two components of skill are not closely related, perhaps because the skill sets that drive each differ considerably. Thus, among fund managers, skillful industry selection often does not coincide with skillful individual stock selection. In fact, 48 percent of our sample funds have industry and stock alpha point estimates of opposite sign.

4. Empirical Analysis

4.1 Persistence

[51] When we use three-digit NAICS codes to define industries rather than two-digit SIC codes, the corresponding four-factor industry-selection skill fraction is 29 percent.

[52] The mean cross-sectional correlation between the stock (industry) component of alpha and total alpha is 0.68 (0.47).

We next examine performance persistence, the ability of funds to maintain their relative performance over time. Numerous papers have examined persistence in overall skill, finding evidence of persistence in risk-adjusted returns including single- and three-factor alphas over one-year intervals (see, for example, Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Brown and Goetzmann (1995), Malkiel (1995), Elton, Gruber, and Blake (1996), and Carhart (1997)) and in four-factor alphas over shorter quarterly horizons (Bollen and Busse (2005)). We explore whether the prior findings are associated with industry-selection skill, individual-stock-selection skill, or both. Most studies also find evidence of persistence specifically in poor performance regardless of the performance measure or measurement interval, which is typically attributed to high expenses. Since we analyze returns gross of expenses, we are unable to shed light on that form of poor performance persistence.

We analyze persistence by sorting into deciles based on performance and then examining the performance of the deciles the following period. We examine persistence in all three estimates of skill: total alpha, industry alpha, and stock alpha. As before, we base our performance estimates on the gross returns imputed from the portfolio holdings. Our total alpha analysis merely repeats that of earlier papers using our specific sample.

Although previous studies analyze one-year post-ranking horizons most often (e.g., Carhart (1997)), evidence of persistence beyond that attributable to momentum is associated with shorter post-ranking horizons (see Bollen and Busse (2005)). Other studies examine longer, three-year post-ranking horizons (e.g., Gruber (1996)). We examine persistence across several post-ranking horizons ranging from one quarter to three years. Regardless of the post-ranking horizon, we use a one-quarter ranking period, which is consistent with the quarterly snapshot frequency of our holdings data. Thus, we sort funds into deciles based on their performance estimate over a quarterly ranking period and then compute the mean performance estimate of each decile over the subsequent post-ranking period (ranging from one quarter to three years). We assess persistence by comparing the post-ranking performance of the best and worst ranking-period deciles and via the Spearman rank correlation coefficient, which we measure between the ranking-period performance decile and the post-ranking-period performance decile.

Table 16 shows the persistence results. Panel A reports the results for total alpha, and panels B and C report the results for the two components of total alpha, industry alpha (Panel B), and stock alpha (Panel C). Each panel reports the results using four factors in the base regression model. Results based on one- and three-factor regression models are qualitatively similar. The table reports persistence results for select post-ranking horizons ranging from one quarter to three years. The table shows the mean post-ranking performance estimate for each decile, the difference between the mean of decile 10 and the mean of decile 1, and the Spearman rank correlation coefficient that assesses the relation between ranking period and post-ranking-period performance deciles. Note that the table reports mean daily percentage returns associated with each post-ranking quarter, rather than the cumulative returns through each quarter.

The results in Panel A are consistent with results documented elsewhere in the literature. For the shorter post-ranking horizons, the results strongly suggest that performance persists across adjacent time periods. The Spearman rank correlation coefficients are statistically significant at the five percent level or better for post-ranking horizons up to eight quarters. Although the Table does not report the statistic significance of the individual deciles, top decile alphas are statistically significant at the one percent level across all 12 post-ranking quarters.[53] These results are consistent with the persistence results of Bollen and Busse (2005), who also find evidence of short-term persistence after controlling for momentum.

Table 16, Panel B reports persistence results for industry-selection skill. The results in Panel B suggest that industry-selection skill persists, even at longer post-ranking horizons. The Spearman rank correlation coefficient is statistically significant at the one percent level for all post-ranking period horizons, including three years, which is at least a full year beyond the total alpha persistence shown in Panel A. Here, persistence in relative performance is somewhat stronger than persistence in absolute performance, as the top decile alphas attain statistical significance at the ten percent level only through the first post-ranking year (significance levels not shown in the table). These results indicate that managers who allocate their assets to the better industries during one quarter continue to do so through three years.

Panel C of Table 16 reports results that assess persistence in individual-stock-selection skill. Compared to panels A and B, the results in this panel show weaker evidence of a monotonic relation between past performance and future performance. The Spearman correlation coefficient is statistically significant for post-ranking horizons only up to four quarters, compared to 12 quarters for industry-selection skill. Regardless of the length of the post-ranking horizon, individual-stock selection skill shows lower

[53] One needs to interpret the significance of Spearman correlation coefficients with caution. The significance is calculated under the assumption that the sample is random. The significance levels do not incorporate cross-sectional correlations.

correlation between past and future performance and a smaller difference between top and bottom decile post-ranking performance than does industry-selection skill. The top decile by itself fares somewhat better, as its stock alpha statistically significantly differs from zero at the five percent level through the second post-ranking year and at the ten percent level for the third year (not shown in the table). These results suggest that the evidence of persistence in relative performance documented in Panel A is driven by industry selection.[54]

Recall that we construct fund industry returns from individual stock holdings, which leads to performance estimates based on gross mutual fund returns. This approach leaves open the possibility that the top deciles in our persistence sorts charge higher expenses or incur greater transaction costs, such that net relative fund performance would not persist. To explore this possibility, we examine the difference between gross and net fund returns for each of the deciles in Table 3. We take net fund returns from the CRSP Survivors-Bias Free U.S. Mutual Fund database and use MFLINKS from WRDS to connect the data to Thomson Financial's portfolio holdings. The gross-net difference represents a combination of expense ratio, transaction costs, the return effects of cash and other non-domestic equity holdings, and performance associated with intra-period trading.

In non-tabulated results, we find that the best performing deciles have slightly higher expenses and/or transaction costs than other deciles, but that the higher costs account for only a small fraction of the difference in industry alpha between decile 10 and decile 1. In particular, the gross-net difference for decile 10 (1) is 0.0077 (0.0066)

[54] When we use three-digit NAICS codes to define industries, the persistence results are very similar to the results in Table 3 based on two-digit SIC codes.

percent per day, or 1.96 (1.68) percent per year. The 0.28 percent annual difference represents approximately one seventh of the 2.06 percent annual difference in gross industry alpha during the first post-ranking quarter and about one fifth of the difference during the fourth post-ranking quarter. These results therefore suggest that the top decile funds would also outperform the bottom decile funds net of expenses and transaction costs.

Given the industry momentum results of Moskowitz and Grinblatt (1999), we might expect lucky funds to continue to outperform on the basis of industry skill to the extent that they do not substantially change their industry allocation over time. To explore the importance of luck in our industry persistence results, we compute industry turnover for each fund, and specifically examine the extent to which funds with superior industry skill change their industry composition over time. Luck is less likely to play a role in our results if top funds substantially change their industry allocation over time, since funds are unlikely to re-allocate to stocks in top-performing industries by chance over long time frames. Recall that we detected industry persistence in Table 3 through 12 post-ranking quarters.

We define quarterly industry turnover as the fraction of a fund's original industry allocation that changes from one quarterly holdings snapshot to the next. We find that the top decile of funds in Table 16, Panel B turns over its industry allocation at a rate of 23 per quarter. Industry turnover across all deciles averages 18 percent. When annualized, these quarterly turnover rates are roughly in line with annual estimates of stock turnover that are also derived from quarterly holdings data (for example, Wermers (2000) reports annual stock turnover of 70 percent over a common sample period). Note that these turnover calculations under-estimate actual turnover, since they do not incorporate intra-quarter portfolio transactions. The industry turnover estimates suggest that persistence in industry selection skill is not likely to be attributable to luck, since it is difficult to imagine how funds that change their entire industry allocation roughly once per year, on average, could persistently invest in better-performing industries by chance over a multi-year time frame.

We next examine the extent to which total alpha, industry alpha, and stock alpha predict future total alpha. To do so, we use a different methodology. Each quarter, we regress cross sectionally future performance on past performance:

$$\alpha_{p,t} = a + b\alpha_{p,t-1} + \varepsilon_{p,t}, \tag{4}$$

where α_p is from regression equation (1). We then compute the mean of the regression coefficients across quarters and compute Fama MacBeth (1973) *t*-statistics. A significant positive *b* coefficient would be consistent with predictability. We repeat the cross-sectional regressions in equation (4) except replacing the α_p regressor first with α_{pi},

$$\alpha_{p,t} = a + b\alpha_{pi,t-1} + \varepsilon_{p,t}, \tag{5}$$

then with α_{ps},

$$\alpha_{p,t} = a + b\alpha_{ps,t-1} + \varepsilon_{p,t}, \tag{6}$$

and finally with both α_{pi} and α_{ps} simultaneously,

$$\alpha_{p,t} = a + b\alpha_{pi,t-1} + c\alpha_{ps,t-1} + \varepsilon_{p,t}. \tag{7}$$

The cross-sectional regression methodology allows us to jointly assess the importance of past industry and stock alphas in predicting future total alpha. We repeat our analysis for alphas based on all three sets of factors.

Table 17 reports the cross-sectional regression results. The table reports the mean *b* and *c* coefficients (averaged across the quarterly regressions), Fama MacBeth (1973) *t*-statistics, and mean r-squares. The left, center, and right sets of columns in the table report the results associated with the single-, three-, and four-factor models, respectively.

Similar to the decile results, the results here again reflect persistence in total alpha. For all three factor models, a statistically significant relation exists between past and future total alpha (regression equation (4)) and between past industry alpha and future total alpha (regression equation (5)). By contrast, no consistent significant relation exists between past stock alpha and future total alpha (regression equation (6)). The coefficient on past stock alpha is statistically significant only for the three-factor model. This result makes sense given that we found only weak evidence of persistence across the stock alpha deciles in Table 3. That is, if past stock alpha does not predict future stock alpha so well, then we would not expect it to predict future total alpha.

The last set of results in Table 4 jointly examines the relation between future total alpha and past industry and stock alpha, as in regression equation (7). The results confirm that persistence in total alpha is driven by the industry component of alpha, rather than the stock component.[55] For the single- and four-factor models, future total alpha is positively and statistically significantly related to past industry alpha, but insignificantly

[55] The results might be misleading because there may be more of a measurement error problem for stock alphas, as compared to using industry alphas. To alleviate this concern, we delete funds with less 20 stocks. The results are only slightly altered.

related to past stock alpha. For the three-factor model, future total alpha is positively and statistically significantly related to both past industry alpha and past stock alpha.[56]

To determine whether the advantages associated with industry-selection skill are economically important, we repeat the analysis in Table 16 except that we sort into deciles based on industry selection skill or stock selection skill, and then examine the mean post-ranking total alphas of each decile. Table 18 shows the results. The table shows that the post-ranking total alpha of the top industry-selection skill decile is greater than the top stock-selection skill decile during all post-ranking quarters. Similarly, post-ranking total alpha differences between the top and bottom deciles are always greater for the industry-selection skill sorts than for the stock-selection skill sorts during all post-ranking quarters. Although the post-ranking top decile total alpha for the industry alpha sort does not statistically significantly differ from that for the stock alpha sort, the 10-1 differences of the industry alpha sort are statistically significantly greater than those of the stock alpha sort at the ten percent level for post-ranking quarters 1-3. This result provides further evidence that industry allocations play an important role in the persistence of relative performance. Overall, the results in Tables 16, 17, and 18 suggest that industry selection, rather than stock selection, drives the evidence of overall performance persistence.

4.3 Stock and Industry Selectivity vs. Fund Size

Ippolito (1992), Gruber (1996), and Sirri and Tufano (1998) show that investors chase winners, since cash flows correlate positively with past performance. As their asset

[56] Although not shown in Table 4, the persistence results are very similar when we use three-digit NAICS codes rather than two-digit SIC codes to define industries, with significant positive relations between future total alpha and past industry alpha and weaker relations between future total alpha and past individual-stock-selection alpha.

bases swell, top-performing funds find it increasingly difficult to maintain stellar performance. Popular funds experience diseconomies of scale, as indicated in Berk and Green (2004) and Chen et al. (2004). One key implication from Berk and Green's (2004) theoretical model is that top performance should not persist indefinitely, as increasing transaction costs associated with larger transactions (e.g., price pressure), the exhausting of one's preferred stock list, or greater fees charged by the managers work to eliminate excess performance. Similarly, poor performance could reverse. For instance, if their asset bases shrink, lagging funds may find it easier to manage their remaining assets, perhaps because they can focus on their best ideas.

In this section, we examine the effects of fund size on the industry-selection and stock-selection components of mutual fund performance. Ex ante, reasons exist to believe that the size of a fund's asset base could differentially affect the two components of performance. As mentioned above, one of the main contributors to diseconomies of scale is the increase in transaction costs associated with large stock trades. That is, if a 100-stock fund grows its asset base ten-fold, but continues holding the same 100 stocks, the fund will need to trade ten times as many shares per stock. What previously could be accomplished with a 1,000-share trade would now require a 10,000-share trade. For all but the most liquid stocks, transacting substantially larger quantities is considerably more difficult, as market impact tends to move prices in the wrong direction. To avoid larger per-share transaction costs, funds may eliminate from consideration stocks that lack sufficient liquidity. Alternatively, funds may choose to increase the number of stocks to hold in their portfolios, an effect consistent with the portfolio data in Table 1. However, since their favorite stocks are typically already in their portfolio, the new additions could hurt fund performance. That is, they almost certainly are less optimistic about the new additions, or they would have already had them in their portfolio. Alexander, Cici, and Gibson (2007), for example, find that the stocks that funds purchase in order to absorb excess cash underperform their valuation-motivated purchases. So, unless a fund can continue to generate additional stock picks that they like as much as their core holdings, getting larger would be expected to hurt their stock selectivity.

Consider, however, a fund that focuses on maintaining a particular industry allocation. A given industry consists of numerous individual issues, often consisting of an assortment of market capitalizations, share prices, and trading volumes. A manager that finds it too costly to transact too much in one stock could add another stock in the same industry. The fund manager would, thus, have numerous opportunities to maintain a specific industry exposure without having to exert undo pressure on any one particular stock. Alternatively, the fund manager could begin investing in a closely-related industry. Consequently, we might anticipate industry-selection ability to suffer less from a larger base of assets than individual-stock-selection ability.

To examine the relation between fund size and performance, we sort funds into deciles based on the size of their stock portfolios at the beginning of the quarter, and then examine the performance of the portfolios over the course of the quarter. Similar to Chen et al. (2004), we examine total alpha, but we also examine the two distinct components of total alpha, industry and stock alpha. We assess the relation between fund size and performance with the Spearman rank correlation coefficient, measured between the beginning-of-quarter size decile and the subsequent performance decile, and with the difference in mean post-ranking performance for the largest and smallest size deciles.

Table 19 shows the results. Panel A reports the results for total alpha, Panel B reports the results for industry alpha, and Panel C reports the results for stock alpha. The total alpha results in Panel A show a negative correspondence between fund size and total alpha. The negative correspondence is statistically significant for all factor models, although the three-factor results are only marginally significant. Thus, on a total alpha basis, larger funds tend to underperform smaller funds, on average, consistent with the findings of Chen et al. (2004). Note that the negative relation between size and total alpha exists even without accounting for transaction costs (since we use gross fund returns).[57] The picture that emerges from these results and the earlier ones is that good performance generates inflows and a larger base of assets, which eventually leads to a subsequent deterioration in performance, as in Berk and Green (2004). However, the performance hit is not sufficient to eliminate persistence immediately (Table 3, Panel A).

The industry results in Panel B are mixed. For the single-factor model, we see some evidence of diseconomies of scale. By contrast, the three- and four-factor results show no evidence of a deteriorating industry alpha as fund size increases. For these models, the evidence is consistent with the opposite result, with a statistically significantly positive Spearman correlation between fund size and industry alpha. The mean industry alpha of the largest size decile, however, does not statistically significantly differ from that of the smallest-size decile for any of the factor models. Overall, then, no strong relation exists between fund size and industry alpha. It appears that fund managers find ample opportunities either in their current industries or possibly in others to maintain their industry performance even as their asset bases grow. In untabulated results, we find

[57] Stock prices on the date of the portfolio holdings snapshot incorporate some price pressure effects for recent purchases.

that this result is particularly strong (i.e., no relation between fund size and industry skill) for funds that hold stocks in larger industries. The lack of a relation between size and industry alpha helps to explain why the industry component of alpha persists, as indicated in Table 16, Panel B.

In Panel C, the stock alpha results strongly point to diseconomies of scale, with Spearman rank correlations between the size decile and subsequent stock alpha near -0.9 for all three factor models. Furthermore, the mean stock alpha of the largest size decile is statistically significantly less than the mean stock alpha of the smallest size decile at the ten percent level or lower for all three factor models. The evidence of diseconomies of scale in total alpha thus appears to be driven entirely by the stock-selection component of alpha. Although funds appear to maintain an equally attractive industry allocation as their size increases, they apparently have a difficult time adding stocks that do as well as their original choices.[58]

5. Conclusion

Some funds excel at picking individual stocks; others stand out with their industry allocations. We find that both types of skill play an important role in ultimately determining a fund's concurrent risk-adjusted performance. We also find that the industry-selection component of total alpha persists more than the stock-selection component, particularly over longer investment horizons. These results suggest that industry-selection ability drives the evidence of performance persistence documented elsewhere in the literature.

[58] NAICS-based results are very similar to the SIC-based results shown in Table 6, with evidence of negative relations between size and total alpha and between size and stock alpha, and mixed evidence for the relation between size and industry alpha.

Investors chase performance, leading to large inflows at the top-performing funds. We find that larger fund sizes do not erode the industry-selection component of performance, possibly because fund managers have ample room to further add to their current industries, or because they are able to find other industries that are equally attractive. By contrast, we find that stock selectivity suffers as fund size increases, consistent with the total performance results of Chen et al. (2004). This result suggests that diseconomies of scale in mutual funds are specifically attributable to the stock-selection component of performance.

References

Admati, A., and P. Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies 1*, 3-40.

Aktas, N., E. de Bodta, F. Declerck, and H.V. Oppens, 2007, The PIN anomaly around M&A announcements, *Journal of Financial Markets 10*, 169-191.

Alexander, G., G. Cici, and Scott Gibson, 2007, Does motivation matter when assessing trade performance? An analysis of mutual funds, *Review of Financial Studies* 20, 125-150.

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

Amihud, Y., and H. Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics 17*, 223–249.

Aslan, H., D. Easley, S. Hvidkjaer, and M. O'Hara, 2007, Firm characteristics and informed trading: Implications for asset pricing, working paper.

Avramov, D., and T. Chordia, 2006, Asset pricing models and financial market anomalies, *Review of Financial Studies 19*, 1001-1040.

Avramov, D., and R. Wermers, 2006, Investing in mutual funds when returns are predictable, *Journal of Financial Economics* 81, 339-377.

Badrinath, S.G., and S. Wahal, 2002, Momentum Trading by Institutions, *Journal of Finance* 57, 2449-2478.

Baker, M. and J. Wurgler, 2006, Investor Sentiment and the Cross-Section of Stock Returns, *Journal of Finance,* 61, 1645-1680.

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

Barber, B., and T. Odean, 2002, Online investors: Do the slow die first?, *Review of Financial Studies 15*, 455-487.

Barber, B. M., and T. Odean, 2008, All that Glitters: The effect of attention and news on the buying behavior of individual and institutional Investors, *Review of Financial Studies 21*, 785–818.

Barber, B., T. Odean, and N. Zhu, 2009, Do Retail Traders Move Markets, *Review of Financial Studies,* 22, 151-186

Barber, B., T. Odean, and N. Zhu, 2009, Systematic Noise, *Journal of Financial Markets*, 12, 547-569 .

Baberis, N., and A. Shleifer, 2003, Style Investing, *Journal of Financial Economics* 68, 161-199.

Barberis, N., A Shleifer, and J. Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283-317.

Banz, R., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3-18.

Barclay, M., and J. Warner, 1993, Stealth trading and volatility, *Journal of Financial Economics 34*, 281-305.

Battalio, R. H., and R.R. Mendenhall, 2005, Earnings expectations and investor clienteles, *Journal of Financial Economics 77*, 289-319.

Berk, J., and R. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269-1295.

Bernard, V.L., and J.K. Thomas, 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics 13*, 305–340.

Bessembinder, H., and H. Kaufman, 1997, A cross-exchange comparison of execution costs and information flow for NYSE-listed stocks, *Journal of Financial Economics 46*, 293-320.

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

Black, F., 1986, Noise, *Journal of Finance 41*, 529–543.

Bollen, N., and J. Busse, 2005, Short-term persistence in mutual fund performance, *Review of Financial Studies* 18, 569-597.

Botosan, C., 1997, Disclosure Level and the Cost of Equity Capital, *The Accounting Review 72*, 323-349.

Botosan, C., and M. Plumlee, 2002, A re-examination of disclosure level and the expected cost of equity capital, *Journal of Accounting Research 40*, 21-40.

Brennan, M.J., T. Chordia, and A. Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics 49*, 345-373.

Brinson, G., L. R. Hood, and G. Beebower, 1986, Determinants of portfolio performance,"*Financial Analysts Journal* 42 (Jul-Aug), 39-44.

Brown, S., and W. Goetzmann, 1995, Performance persistence, *Journal of Finance* 50, 679-698.

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

Chae, J., 2005, Trading volume, information asymmetry, and timing information, *Journal of Finance 60*, 412-442.

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

Chakravarty, S., V. Panchapagesan, and R. Wood, 2005, Did decimalization hurt institutional investors?, *Journal of Financial Markets* 8, 400-420.

Chen, N. F., 1991, Financial investment opportunities and the macroeconomy, *Journal of Finance 46,* 529–554.

Chen, J., H. Hong, M. Huang, and J. Kubik, 2004, Does fund size erode mutual fund performance? The role of liquidity and organization, *American Economic Review* 94, 1276-1302.

Choi, N., and R. Sias, 2009, Institutional Industry Herding, *Journal of Financial Economics* 94, 469-491.

Chordia, T., R. Roll, and A. Subrahmanyam, 2001, Market liquidity and trading activity, *Journal of Finance 56*, 501–530.

Chordia, T., R. Roll, and A. Subrahmanyam, 2009, Why has trading volume increased, working paper.

Chordia, T., A. Subrahmanyam, and R. Anshuman, 2001, Trading activity and expected stock returns, *Journal of Financial Economics 59*, lead article.

Chordia, T., and A. Subrahmanyam. 2004, Order imbalance and individual stock returns: Theory and evidence, Journal *of Financial Economics 72*, 485–518.

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

Datar, V., N. Naik, and R. Radcliffe, 1998, Liquidity and asset returns: An alternative test, *Journal of Financial Markets 1*, 203–220.

Dennis, D. and J. McConnell, 1986, Corporate mergers and security returns, *Journal of Financial Economics 16*, 143–187.

Duarte, J., and L. Young, 2009, Why is PIN priced? *Journal of Financial Economics*, forthcoming.

Easley, D., N.M. Kiefer, and M. O'Hara, 1997, The information content of the trading process. *Journal of Empirical Finance 4*, 159–186.

Easley, D., N.M. Kiefer, and M. O'Hara, 1997b, One day in the life of a very common stock, *Review of Financial Studies 10*, 805-835.

Easley, D., N.M. Kiefer, M. O'Hara, and J. B. Paperman, 1996, Liquidity, information, and infrequently traded stocks, *Journal of Finance 51*, 1405-1436.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, Is Information Risk a Determinant of Asset Returns? *Journal of Finance 57*, 2185-2221.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2009, Factoring information into returns, *Journal of Financial and Quantitative Analysis*, forthcoming.

Easley, D., and M. O’Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics 19*, 69-90.

Easley, D., and M. O’Hara, 1992, Adverse selection and large trade volume: The implications for market efficiency, *Journal of Financial and Quantitative Analysis 27*, 185-208.

Easley, D., and M. O’Hara, 1992, Time and the process of security price adjustment, *Journal of Finance 47*, 577–606.

Easley, D., and M. O’Hara, 2004, Information and the cost of capital, *Journal of Finance 59*, 1553-1583.

Easley, D., M. O'Hara, and J. Paperman, 1996, Liquidity, information, and less-frequently traded stocks, *Journal of Finance 51*, 1405-1436.

Easley, D., M. O'Hara, and J. Paperman, 1998, Financial analysts and information-based trade, *Journal of Financial Markets 1*, 175-201.

Easley, D., M. O'Hara, and G. Saar, 2001, How stock splits affect trading: A microstructure approach, *Journal of Financial and Quantitative Analysis 36*, 25-51.

Elton, E., M. Gruber, and C. Blake, 1996, The persistence of risk-adjusted mutual fund performance, *Journal of Business* 69, 133-157.

Fama, E., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics 49*, 283-306.

Fama, E.F., and K.R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics 19*, 3-29.

Fama, E.F., and K.R. French, 1992, The cross-section of expected stock returns, *Journal of Finance 47*, 427-465.

Fama, E.F., and K.R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics 33*, 3-56.

Fama, E.F., and K.R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Financial Economics 51*, 55-84.

Fama, E.F., and J.D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy 81*, 607-636.

Foster, F. Douglas, and S. Viswanathan, 1990, A theory of the interday variations in volume, variance, and trading costs in securities markets, *Review of Financial Studies 3*, 593–624.

Foster, F.D., and S. Viswanathan, 1993, The effect of public information and competition on trading volume and price volatility, *Review of Financial Studies 6*, 23-56.

Frazzini, A., and O. Lamont, 2007, The earnings announcement premium and trading volume, working paper.

Frazzini, A., and O. Lamont, 2007, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics*, forthcoming.

Froot, K., D. Scharfstein, and J. Stein, 1992, Herd on the Street: Informational Inefficiencies in a Market with Short-term Speculation, *Journal of Finance* 47, 1461-1484.

Froot. K and M. Teo, 2008, Style Investing and Institutional Investors, *Journal of Financial and Quantitative Analysis* 43, 883-906.

Gallant, R., P. Rossi, and G. Tauchen, 1992, Stock prices and volume, *Review of Financial Studies 5*, 199-242.

Gervais, S., and T. Odean, 2001, Learning to be overconfident, *Review of Financial Studies 14*, 1-27.

Gervais, S., R. Kaniel, and D. Mingelgrin, 2001, The high-volume return premium, *Journal of Finance 56*, 877-919.

Glosten, L. R., and L. Harris, 1988, Estimating the components of the bid-ask spread, *Journal of Financial Economics 21*, 123–42.

Glosten, L., and P. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics 13*, 71-100.

Goetzmann, W., and R. Ibbotson, 1994, Do winners repeat? Patterns in mutual fund performance, *Journal of Portfolio Management* 20, 9-18.

Gomes, J., L. Kogan, and L. Zhang, 2003, Equilibrium cross-section of returns, *Journal of Political Economy 111*, 693-732.

Goyal, A., and I. Welch, 2007, Performance of equity premium prediction, *Review of Financial Studies 21*, 1455-1508.

Green, C., and B. Hwang, 2009, Price-Based Return Comovement, *Journal of Finance Economics* 93, 37-50.

Greenwood, R., 2008, Excess Comovement of Stock Returns: Evidence from Cross-sectional Variation in Nikkei 225 Weights, *Review of Financial Studies* 21, 1153-1186.

Griffin, J., J. Harris, and S. Topaloglu, 2003, The Dynamics of Institutional and Individual Trading, *Journal of Finance* 58, 2285-2320.

Grinblatt, M., and S. Titman, 1989, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of Business* 62, 393-416.

Grinblatt, M., and S. Titman, 1992, The persistence of mutual fund performance, *Journal of Finance* 47, 1977-1984.

Grinblatt, M., S. Titman, and R. Wermers, 1995, Momentum Investing Strategies, Portfolio Performance, and Herding: a Study of Mutual Fund Behavior, *American Economic Review* 85, 1088-1105.

Gruber, M., 1996, Another puzzle: The growth in actively managed mutual funds, *Journal of Finance* 51, 783-810.

Hasbrouck, J., 1991, Measuring the information content of stock trades, *Journal of Finance 46*, 179-207.

Hasbrouck, J., 1988, Trades, quotes, inventories, and information, *Journal of Financial Economics 22*, 229-252.

Hausman, J., A. Lo, and C. MacKinlay, 1992, An ordered probit analysis of transaction stock prices, *Journal of Financial Economics 31*, 319-379.

Hendricks, D., J. Patel, and R. Zeckhauser, 1993, Hot hands in mutual funds: Short-run persistence of performance, 1974-1988, *Journal of Finance* 48, 93-130.

Hwang, C.Y., and Qian, X, 2009, Is information risk priced? Evidence from the price discovery of large trades, Working paper, Nanyang Technological University, Singapore.

Hirshleifer, D., A. Subrahmanyam, and S. Titman, 1994, Security Analysis and Trading Patterns when some Investors Receive Information before Others, *Journal of Finance* 49, 1665-1698.

Hirshleifer, D., J. N. Myers, L. A. Myers, and S. H. Teoh, 2008, Do individual investors drive post-earnings-announcement drift? Direct evidence from personal trades. *Accounting Review*, forthcoming.

Holthausen, R., and Verrecchia, R., 1990, The e!ect of informedness and consensus on price and volume behavior, *Accounting Review 65*, 191-208.

Holden, C.W., and A. Subrahmanyam, 1992, Long-lived private information and imperfect competition, *Journal of Finance 47*, 247-270.

Huang, R., and H. Stoll, 1997, The components of the bid-ask spread: A general approach, *Review of Financial Studies 7*, 995–1034.

Hughes, J., J. Liu, and J. Liu, 2007, Private information, diversification, and asset pricing, *Accounting Review*, forthcoming.

Hvidkjaer, S., 2006, A Trade-Based Analysis of Momentum, *Review of Financial Studies 19*, 457–91.

Hvidkjaer, S., 2008, Small Trades and the Cross-section of Stock Returns, *Review of Financial Studies*, forthcoming.

Ippolito, R., 1992, Consumer reaction to measures of poor quality: Evidence from the mutual fund industry, *Journal of Law and Economics* 35, 45-70.

Jegadeesh, N., 1990, Evidence of predictable behavior in security returns, *Journal of Finance 45*, 881-898.

Jegadeesh, N., and S. Titman 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance 48*, 65-91.

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

Jones, C., Kaul, G., and M. Lipson, 1994, Transactions, volume and volatility, *Review of Financial Studies 7*, 631-652.

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

Kacperczyk, M., C. Sialm, and L. Zheng, 2005, On the industry concentration of actively managed mutual funds, *Journal of Finance* 60, 1983-2011.

Kandel, E., and Pearson, N., 1995, Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy*, 103, 831-72.

Karpoff, J., 1987, The relation between price change and trading volume: A survey, *Journal of Financial and Quantitative Analysis 22*, 109-126.

Keim, D., and R.F. Stambaugh, 1986, Predicting returns in the stocks and the bond markets, *Journal of Financial Economics 17*, 357-390.

Keown, A., and J. Pinkerton, 1981, Merger announcement and insider trading activity: An empirical investigation, *Journal of Finance 36*, 855–869.

Kim, O., and Verrecchia, R., 1991, Market reactions to anticipated announcements, *Journal of Financial Economics 30*, 273-310.

Koski, J.L., and Michaely, R., 2000, Prices, liquidity, and the information content of trades, *Review of Financial Studies 13*, 659-696.

Kumar, A., 2009, Dynamic Style Preferences of Individual Investors and Stock Returns, *Journal of Financial and Quantitative Analysis* 44, 607-640.

Kumar, A and G. Korniotis, 2010, Do Portfolio Distortions Reflect Superior Information or Psychological Biases, working paper, Univeristy of Texas.

Kumar, A., and C. Lee, 2006, Retail Investor Sentiment and Return Comovements, *Journal of Finance* 61, 2451-2486.

Kyle, A. S., 1985, Continuous auctions and insider trading, *Econometrica 53*, 1315-1335.

Lakonishok, J., A. Shleifer, and R. Vishny, 1992, The Impact of Institutional Trading on Stock Prices, *Journal of Financial Economics* 32, 23-43.

Lambert, R., C. Leuz, and R.E. Verrecchia, 2007, Accounting information, disclosure, and the cost of capital, *Journal of Accounting Research 45*, 385–420.

Lee, C.M. 1992, Earnings news and small traders: An intraday analysis, *Journal of Accounting and Economics,* 15, 265-302.

Lee, C.M., and M. Ready, 1991, Inferring trade direction from intradaily data, *Journal of Finance 46*, 733-746.

Lee, C., and B. Swaminathan, 2000, Price momentum and trading volume, *Journal of Finance 55*, 2017-2069.

Lehmann, B., 1990, Fads, Martingales, and Market Efficiency, *Quarterly Journal of Economics* 105, 1-28.

Levi, S., and X.J. Zhang, 2009, Information asymmetry and the increase in expected stock returns before earnings announcements, working paper.

Lin, J., G.C. Sanger, and G.G. Booth, 1995, Trade size and components of the bid-ask spread, *Review of Financial Studies 8*, 1153-1 184.

Linnainmaa, J., 2010, Do Limit Orders Alter Inferences About Investor Performance and Behavior, forthcoming, *Journal of Finance*.

Litzenberger, R., and K. Ramaswamy, 1979, The effect of personal taxes and dividends on capital asset pricing: Theory and empirical evidence, *Journal of Financial Economics 7*, 163-196.

Lo, A., and C. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies 3*, 431-467.

Malkiel, B., 1995, Returns from investing in equity mutual funds 1971 to 1991, *Journal of Finance* 50, 549-572.

Merton, R. C., 1987, A simple model of capital market equilibrium with incomplete information, *Journal of Finance 42*, 483-510.

Meulbroek, L., 1992, An empirical analysis of illegal insider trading, *Journal of Finance 47*, 1661–1699.

Mohanram, P., and S. Rajgopal, 2008, Is PIN priced risk?, *Journal of Accounting and Economics*, forthcoming.

Moskowitz, T., and M. Grinblatt, 1999, Do industries explain momentum?, *Journal of Finance* 54, 1249-1290.

Newey, W., and K. West, 1987, A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica 55*, 703-708.

Newey, W., and K. West, 1994, Automatic lag selection in covariance matrix estimation, *Review of Economic Studies 61*, 631-653.

Odean, T., 1998, Volume, volatility, price, and profit when all traders are above average, *Journal of Finance 53*, 1887-1934.

O'Hara, M., 2003, Presidential Address: Liquidity and price discovery, *Journal of Finance 58*, 1335-1354.

Pfleiderer, P., 1984, The volume of trade and variability of prices: A framework for analysis in noisy rational expectations equilibria, working paper, Stanford University.

Seppi, D., 1990, Equilibrium block trading and asymmetric information, *Journal of Finance 45,* 73-94.

Seppi, D., 1992, Block trading and information revelation around quarterly earnings announcements, *Review of Financial Studies 5*, 281-305.

Shanken, J., 1990, Intertemporal Asset Pricing: An Empirical Investigation, *Journal of Econometrics 45*, 99-120.

Shanken, J., 1992, On the estimation of beta-pricing models, *Review of Financial Studies 5*, 1-33.

Sias, R., 2004, Institutional Herding, *Review of Financial Studies* 17, 165-206.

Sias, R., L. Starks, and S. Titman, 2006, Changes in Institutional Ownership and Stock Returns: Assessment and Methodology, *Journal of Business* 79, 2869-2910.

Sirri, E., and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589-1622.

Statman, M, S. Thorley, and K. Vorkink, 2006, Investor overconfidence and trading volume, *Review of Financial Studies 19*, 1531-65.

Stoll, H.R., 1978, The supply of dealer services in securities markets, *Journal of Finance 33*, 1133-1151.

Teo, M., and S. Woo, 2004, Style Effects in the Cross Section of Stock Returns, *Journal of Financial Economics* 74, 367-398.

Vega, C., 2006, Stock price reaction to public and private information, *Journal of Financial Economics 82*, 103-133.

Wang, J., 1993, A model of intertemporal asset prices under asymmetric information, *Review of Economic Studies 60*, 249-282.

Wang, J., 1994, A model of competitive stock trading volume, *Journal of Political Finance 102*, 127–168.

Wermers, R., 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses, *Journal of Finance* 55, 1655-1695.

Wolff, E., 2004, Changes in Household Wealth in the 1980s and 1990s in the US, working paper, The Levy Economics Institute of Bard College.

Ying, C., 1966, Stock market prices and volumes of sales, *Econometrica 34*, 676-685.

Zheng, L., 1999, Is money smart? A study of mutual fund investors' fund selection ability, *Journal of Finance* 54, 901-933.

Table 1
Summary statistics

Panel A reports average percentage of number of trades and percentage of dollar trading volume by each type of trade. Panel B shows mean, median and standard deviation of changes in total trading volume (AT), small trading volume (AS), two medium size trading volume (AMT1 and AMT2) and large trading volume (ALT). The sample period is from 1983 to 2006.

Panel A: Percentage of small, medium size1, medium size2 and large trades

	large trade	Medium-size2 trade	Medium-size1 trade	Small trade
% of number of trades	9.7%	30.2%	30.9%	29.2%
% of dollar volume	33.2%	36.2%	25.3%	5.3%

Panel B: Summary statistics of AT, ALT, AMT2, AMT1, AST

	AT	ALT	AMT2	AMT1	AST
Mean	0.068	0.066	0.073	0.075	0.086
Median	0.061	0.060	0.065	0.064	0.075
Std	0.645	0.712	0.584	0.610	0.608

Table 2
Correlation matrix

The table contains the time-series means of monthly bivariate correlations of the variables. ALT, AMT2, AMT1, AST and AT refer to changes in large trades, two medium-size group trades, small trades and total trades, respectively. Size represents the logarithm of market capitalization in billions of dollars. BM is the logarithm of the book-to-market ratio. TURN represents logarithm of share turnover. ILLIQ is the Amihud (2002) illiquidity measure. SPREAD represents the quoted spread. PIN is a measure of probability of information-related trading estimated with the Easley, Hvidkjaer and O'Hara (2002) trading model.

	ALT	AMT2	AMT1	AST	AT	SIZE	BM	TURN	ILLIQ	SPREAD	PIN
ALT	1	0.274	0.104	0.045	0.281	0.113	-0.168	0.283	-0.042	0.071	0.052
AMT2		1	0.316	0.171	0.208	0.148	-0.194	0.200	-0.040	0.032	-0.090
AMT1			1	0.316	0.104	0.144	-0.174	0.156	-0.035	0.002	-0.092
AST				1	0.274	0.148	-0.194	0.200	-0.038	-0.027	-0.137
AT					1	0.113	-0.168	0.283	-0.039	0.030	-0.058
SIZE						1	-0.379	0.392	-0.201	-0.487	-0.621
BM							1	-0.195	0.109	0.182	0.275
TURN								1	-0.249	-0.289	-0.209
ILLIQ									1	0.048	0.218
SPREAD										1	0.191
PIN											1

Table 3
Returns to abnormal trading volume portfolios

Quintiles are formed monthly based on changes in large trading volume (ALT, Panel A), medium size 2 trading volume (AMT2, Panel B), medium size 1 trading volume (AMT1, Panel C), small trading volume (AST, Panel D) in the previous month. Stocks with low (high) changes are in quintile 1 (5). The table reports equally weight gross returns, as well as risk-adjusted returns (alpha) using the CAPM and Fama-French three factors. The difference in returns between the high and the low changes are also reported. The sample period is from 1983 to 2006.

	% Gross returns	% Alphas (CAPM)	% Alphas (FF)
		Panel A: ALT portfolios	
1	0.80	-0.35	-0.50
2	1.03	-0.02	-0.13
3	1.09	0.04	-0.08
4	1.20	0.13	0.02
5	1.40	0.27	0.15
5-1	0.60(3.20)	0.62(3.15)	0.65(3.36)
		Panel B: AMT2 portfolios	
1	0.96	-0.15	-0.27
2	1.15	0.07	-0.05
3	1.11	0.02	-0.08
4	1.12	0.04	-0.11
5	1.20	0.09	0.01
5-1	0.25(1.32)	0.24(1.27)	0.28(1.49)
		Panel C: AMT1 portfolios	
1	1.05	-0.03	-0.16
2	1.03	-0.10	-0.21
3	1.11	0.02	-0.09
4	1.19	0.10	-0.02
5	1.17	0.08	-0.04
5-1	0.12(0.71)	0.11(0.66)	0.12(0.65)
		Panel D: AST portfolios	
1	0.98	-0.14	-0.28
2	1.04	-0.02	-0.11
3	1.20	0.11	0.00
4	1.13	0.05	-0.09
5	1.18	0.07	-0.06
5-1	0.19(0.85)	0.21(0.92)	0.22(1.07)

Table 4
Returns to double-sort portfolios

Five portfolios are formed monthly based on market capitalization (Panel A) or changes in total trading volume (Panel B), and within each portfolio, quintiles are formed based on ALT in the previous month. "1" ("5") represents the low (high) value. The table reports equally weighted gross returns, as well as risk-adjusted returns (alpha) using Fama-French three factors. The difference in returns between the high and the low sell lambda portfolios are also reported.

	% Gross returns						% FF alpha					
	Panel A: Sort by size, then ALT											
	ALT						ALT					
	1	2	3	4	5	5-1	1	2	3	4	5	5-1
Size 1	0.81	1.12	1.33	1.35	1.59	0.78	-0.79	-0.30	-0.26	0.14	0.05	0.84
						(4.12)						(4.97)
2	0.90	0.97	1.48	1.52	1.61	0.71	-0.61	-0.38	-0.31	0.03	0.20	0.81
						(3.61)						(3.53)
3	0.89	1.01	1.31	1.39	1.45	0.56	-0.33	-0.16	0.06	0.11	0.12	0.45
						(2.58)						(2.15)
4	0.70	0.86	1.09	0.91	1.16	0.46	-0.32	-0.36	0.08	0.00	0.08	0.40
						(2.78)						(2.02)
Size 5	0.63	0.85	0.91	1.08	0.99	0.36	-0.13	-0.09	0.11	0.25	0.25	0.38
						(1.85)						(1.90)
	Panel B: Sort by abnormal total trading volume, then ALT											
	ALT						ALT					
	1	2	3	4	5	5-1	1	2	3	4	5	5-1
AT 1	0.60	0.73	0.75	0.98	1.03	0.43	-0.55	-0.36	-0.50	-0.16	-0.14	0.41
						(1.99)						(2.39)
2	0.75	0.90	0.95	1.07	1.22	0.47	-0.42	-0.17	-0.26	-0.04	-0.03	0.39
						(2.87)						(2.12)
3	0.87	1.34	1.29	1.39	1.20	0.33	-0.24	0.18	0.09	0.13	0.11	0.36
						(1.83)						(1.92)
4	0.97	1.30	1.33	1.36	1.37	0.41	-0.30	0.04	-0.04	0.09	0.16	0.46
						(2.37)						(2.80)
AT 5	0.89	1.09	1.20	1.35	1.40	0.51	-0.35	-0.08	0.04	0.12	0.07	0.42
						(3.29)						(2.68)

Table 5

Fama-MacBeth regression estimates with excess market return, SMB and HML as risk factors

This table presents the time-series averages of individual stock cross-sectional OLS regression coefficient estimates. “Unscaled” columns represent that the dependent variable is the excess return risk-adjusted using the Fama-French (1993) factors. “Size+BM” columns represent that the dependent variable is the excess return risk-adjusted using the Fama-French (1993) factors with loadings scaled by size and book-to-market ratio. “Term+Def+Tbill” columns represent that the term spread, the default spread and the T-bill yield are used as scaling variables. Size represents the logarithm of market capitalization in billions of dollars. BM is the logarithm of the book-to-market ratio with the exception that book-to-market ratios greater than the 0.995 fractile or less than the 0.005 fractile are set equal to the 0.995 and the 0.005 fractile values, respectively. TURN represents the logarithm of turnover. CVTURN is a proxy for the volatility of turnover. RET2-12 are the cumulative returns over the second through twelfth months prior to the current month. In columns [1], AT is used as additional independent variable; in columns [2] ALT is included; in columns [3] both AT and ALT are included, in columns [4] ALT, AMT2, AMT1 and AST are included as independent variables. t-statistics in parenthesis use standard errors as per Shanken (1992). All coefficients are multiplied by 100.

	[1] Unscaled	[2] Unscaled	[3] Unscaled	[4] Unscaled	[1] Size+BM	[2] Size+BM	[3] Size+BM	[4] Size+BM	[1] Term+Def+ Tbill	[2] Term+Def+ Tbill	[3] Term+Def+ Tbill	[4] Term+Def+ Tbill
Intercept	-0.150	-0.142	-0.138	-0.129	-0.167	-0.157	-0.159	-0.170	-0.128	-0.131	-0.129	-0.126
	(-0.23)	(-0.40)	(-0.40)	(-0.42)	(-0.73)	(-0.47)	(-0.42)	(-0.47)	(-0.23)	(-0.24)	(-0.27)	(-0.27)
SIZE	0.036	0.038	0.033	0.045	0.020	0.028	0.023	0.026	0.036	0.037	0.037	0.038
	(0.58)	(0.69)	(0.71)	(0.89)	(0.28)	(0.62)	(0.75)	(0.62)	(0.51)	(0.69)	(0.78)	(0.80)
BM	0.175	0.182	0.180	0.171	0.177	0.187	0.181	0.167	0.183	0.182	0.181	0.173
	(3.71)	(3.79)	(3.82)	(3.72)	(3.80)	(3.80)	(3.75)	(3.87)	(3.65)	(3.80)	(3.70)	(3.65)
TURN	-0.261	-0.270	-0.276	-0.269	-0.291	-0.286	-0.289	-0.285	-0.273	-0.277	-0.271	-0.282
	(-5.69)	(-5.89)	(-5.56)	(-5.32)	(-5.43)	(-5.40)	(-5.42)	(-5.41)	(-5.36)	(-5.51)	(-5.29)	(-5.59)
CVTURN	-0.432	-0.427	-0.412	-0.423	-0.453	-0.457	-0.461	-0.451	-0.442	-0.448	-0.441	-0.446
	(-4.72)	(-4.68)	(-4.62)	(-4.67)	(-5.02)	(-5.07)	(-5.05)	(-4.98)	(-4.86)	(-4.92)	(-4.81)	(-4.87)
RET2-12	1.148	1.135	1.137	1.142	1.132	1.133	1.128	1.127	1.156	1.160	1.152	1.157
	(3.03)	(2.99)	(2.91)	(3.05)	(2.91)	(2.89)	(2.82)	(2.86)	(3.11)	(3.07)	(3.05)	(3.12)
AT	0.160		0.116		0.165		0.118		0.170		0.111	
	(2.07)		(1.53)		(2.17)		(1.58)		(2.08)		(1.52)	
ALT		0.334	0.285	0.306		0.318	0.273	0.311		0.312	0.275	0.291
		(4.02)	(3.79)	(4.09)		(3.98)	(3.66)	(4.12)		(3.99)	(3.69)	(4.06)
AMT2				0.017				0.014				0.021
				(0.82)				(0.79)				(0.88)
AMT1				0.063				0.069				0.065
				(1.35)				(1.50)				(1.35)
AST				-0.030				-0.032				-0.030
				(-0.80)				(-0.88)				(-0.85)

Table 6
ALT and PIN

This table presents the time-series averages of individual stock cross-sectional OLS regression coefficient estimates. The dependent variable is the excess return risk-adjusted using the Fama-French (1993) factors. SIZE represents the logarithm of market capitalization in billions of dollars. BM is the logarithm of the book-to-market ratio with the exception that book-to-market ratios greater than the 0.995 fractile or less than the 0.005 fractile are set equal to the 0.995 and the 0.005 fractile values, respectively. TURN represents turnover. RET2-12 is the cumulative returns over the second through twelfth months prior to the current month. ALT is the change in large trades. ILLIQ is the Amihud (2002) illiquidity measure. PIN is a measure of probability of information-related trading estimated with the Easley, Hvidkjaer, and O'hara (2002) trading model. PSOS is a measure of illiquidity unrelated to information asymmetry, as described in Duarte and Young (2009). The t-statistics in parenthesis use standard errors as per Shanken (1992). All coefficients are multiplied by 100.

	[1]	[2]	[3]	[4]	[5]	[6]
Intercept	-0.161	-0.126	-0.112	-0.128	-0.120	-0.130
	(-0.73)	(-0.40)	(-0.29)	(-0.43)	(-0.33)	(-0.36)
SIZE	0.011	0.041	0.047	0.043	0.027	0.035
	(0.11)	(0.81)	(1.03)	(0.85)	(0.59)	(0.68)
BM	0.086	0.174	0.165	0.179	0.098	0.180
	(2.01)	(3.76)	(3.71)	(3.70)	(2.45)	(3.75)
TURN	-0.162	-0.256	-0.243	-0.250	-0.189	-0.229
	(-3.28)	(-5.21)	(-5.01)	(-5.13)	(-4.11)	(-5.12)
RET2-12				1.118		1.142
				(2.78)		(2.98)
ALT					0.278	0.262
					(3.02)	(2.79)
PIN	0.982	0.390	0.422	0.206	0.589	0.170
	(1.87)	(0.60)	(0.92)	(0.35)	(1.45)	(0.16)
PSOS			0.321			0.293
			(1.36)			(1.21)
ILLIQ		0.423				0.418
		(1.78)				(1.73)

Table 7
Abnormal trading volume around different corporate events

This table contains the daily abnormal trading volume around different types of announcements from the companies in the NYSE and the AMEX between 1983 and 2006. Small/medium1/medium2/large refer to abnormal trading volume for each type of trade (see equation (10)). Dif is the difference between small and large abnormal trading volume. (-10,- 4), -3, -2, -1, 0, (1, 2), (3, 10) are the average abnormal trading volume over t=[-10,-4], -3, -2, -1, 0, [1, 2] and [3,10], respectively. Numbers are multiplied by 100. The t-statistics computed using the White correction for heteroskedasticity are reported in parenthesis.

	Earnings					Acquisition					Target				
	Small	M1	M2	Large	Dif	Small	M1	M2	Large	Dif	Small	M1	M2	Large	Dif
(-10, -4)	-1.12	-1.01	-0.67	-0.58	0.70	0.32	0.38	0.35	0.57	0.25	0.69	0.70	0.90	1.71	1.02
	(-3.56)	(-4.61)	(-3.50)	(-2.96)	(3.19)	(1.40)	(2.11)	(2.30)	(2.96)	(2.01)	(2.59)	(2.91)	(2.76)	(3.79)	(3.50)
-3	1.03	1.53	1.70	2.12	1.09	2.53	2.97	3.18	3.34	0.81	2.28	2.33	3.32	3.96	1.68
	(2.58)	(3.40)	(3.88)	(4.82)	(3.95)	(5.34)	(5.98)	(6.29)	(6.54)	(3.87)	(5.86)	(5.34)	(7.98)	(8.92)	(4.34)
-2	3.02	3.56	3.87	4.70	1.68	4.76	5.55	6.54	8.61	3.85	8.50	10.11	9.39	12.53	4.03
	(7.03)	(8.45)	(7.75)	(10.35)	(4.02)	(9.50)	(11.32)	(15.19)	(14.56)	(11.70)	(10.31)	(18.77)	(21.23)	(26.50)	(12.84)
-1	3.76	4.32	4.49	5.76	2.00	6.20	6.77	8.06	11.37	5.17	10.16	15.71	13.29	16.15	5.99
	(7.98)	(9.30)	(9.73)	(12.32)	(5.09)	(11.65)	(14.50)	(17.30)	(18.31)	(13.98)	(14.33)	(22.65)	(26.34)	(33.10)	(16.59)
0	30.29	25.23	28.34	27.11	-3.18	26.67	20.78	27.30	30.45	3.78	45.93	32.53	36.30	42.29	-3.64
	(54.98)	(45.98)	(47.49)	(49.20)	(-9.24)	(40.71)	(43.21)	(56.32)	(48.53)	(6.65)	(87.42)	(70.76)	(50.42)	(69.52)	(-8.97)
(1, 2)	19.33	17.90	12.91	16.33	-3.00	18.40	12.54	20.34	21.34	2.52	29.88	28.12	21.36	24.20	-5.68
	(31.30)	(18.90)	(25.69)	(19.38)	(-8.10)	(23.10)	(31.91)	(39.02)	(30.29)	(4.20)	(50.32)	(45.53)	(32.97)	(49.08)	(-11.09)
(3, 10)	4.22	2.12	1.76	1.27	-2.95	3.11	1.13	0.92	0.75	-2.36	8.30	4.34	2.98	2.39	-5.91
	(10.89)	(5.69)	(7.45)	(3.23)	(-5.43)	(5.12)	(2.10)	(3.21)	(2.60)	(-3.78)	(15.90)	(8.11)	(3.99)	(3.10)	(-11.23)

Table 8

Industry Trading Descriptive Statistics

Each month, from January 1983 to December 2000, we classify stocks into one of the Fama and French (1997) 49 industries. Panel A reports the time-series average of the cross-sectional descriptive statistics for the number of firms in each industry, the percentage of total market capitalization accounted for by each industry, and the fraction of industry capitalization accounted for by the largest firm in the industry. Panel B provides reports the time-series average of the cross-sectional descriptive statistics for number of small and large trades made in each industry, the proportion bought by small and large traders in each industry, and the percentage of total small and large trader industry trading accounted for by the largest firm in the industry.

	Mean	Median	Minimum	Maximum	Std Dev
Panel A: Industry Statistics					
No. of firms in industry	98	61	5	526	83
Industry capitalization/Market capitalization	2.04%	1.32%	0.08%	10.78%	2.17%
Largest firm in industry/Industry capitalization	30.56%	23.21%	4.98%	78.23%	9.34%
Panel B: Industry Trading Statistics					
No. of Small Trades in an Industry	58,456	51,327	3,278	321,243	48,239
No. of Large Trades in an Industry	47,987	42,340	2,861	265,397	37,309
Small trades proportion bought	51.06%	51.02%	36.55%	64.61%	3.03%
Large trades proportion bought	52.72%	52.89%	42.97%	59.89%	2.25%
No. of Small Trades Largest Firm/ No. of Small Trades Industry	27.45%	20.12%	2.87%	69.23%	12.10%
No. of Large Trades Largest Firm/ No. of Large Trades Industry	23.23%	18.13%	3.85%	65.45%	11.65%

Table 9

Firm Herding, Size and Book-to-Market Herding, and Industry Herding

Each year, from 1983 to 2000, all stocks are assigned to one of 49 Fama and French (1997) industry portfolios and one of 6 Fama and French (1993) size and book-to-market portfolios. Each week we run cross-sectional regressions, where the dependent variable is the firm-level proportion bought and the independent variables include lagged firm-level proportion bought, lagged proportion bought from all other stocks in the same industry and the same size and book-to-market portfolio, and lagged proportion bought from all other stocks in the same industry but in a different size and book-to-market portfolio. Proportion bought is lagged 1 week, 2 to 8 weeks, 9 to 24 weeks, 25 to 52 weeks, and 53 to 104 weeks. Panel A reports the results based on retail investor (small trade) proportion bought and Panel B reports the results based on institutional (large trade) proportion bought. The coefficients reported are the time-series averages of the cross-sectional estimates. The standard errors are computed using the Newey-West adjustment. T-statistics are in parentheses.

	Weeks				
	-1	-2 to -8	-9 to -24	-25 to -52	-53 to -104
Panel A: Retail Investors					
Firm Proportion Bought	0.28	0.22	0.06	0.03	0.01
	[28.45]	[17.19]	[9.11]	[5.55]	[1.24]
Size and BM Proportion Bought	0.18	0.15	0.05	0.02	0.01
	[15.34]	[11.22]	[5.39]	[3.89]	[0.98]
Industry Proportion Bought	0.16	0.14	0.05	0.03	0.01
	[19.32]	[14.78]	[7.65]	[4.07]	[0.28]
Adjusted R^2	0.10	0.06	0.04	0.03	0.03
Panel B: Institutional Investors					
Firm Proportion Bought	0.138	0.128	0.027	0.005	0.001
	[14.08]	[5.19]	[2.11]	[1.11]	[0.11]
Size and BM Proportion Bought	0.082	0.054	0.012	0.003	0.001
	[11.39]	[4.21]	[1.78]	[0.19]	[0.78]
Industry Proportion Bought	0.072	0.042	0.008	0.001	0.002
	[10.77]	[3.79]	[1.45]	[0.32]	[0.39]
Adjusted R^2	0.06	0.03	0.02	0.01	0.01

Table 10
Proportion Bought and Prior Industry Returns

This table presents the results from industry-level (Panel A) and firm-level (Panel B) Fama-Macbeth cross-sectional regression estimated monthly from January 1983 to December, 2000. In Panel A, retail investor (institutional) industry proportion bought are regressed on lagged industry returns, lagged retail investor (institutional) industry proportion bought, industry average values of ln (Size) and industry average values of ln(BM). In Panel B, retail investor (institutional) firm-level proportion bought are regressed on lagged industry returns, lagged retail investor (institutional) firm proportion bought, lagged firm returns, firm ln(size) and firm ln(bm). Time-series average values of the monthly regression coefficients are reported blow. Standard errors are adjusted using the Newey-West correction. T-statistics are in brackets.

Panel A: Industry Proportion Bought

	Retail Investor Industry Proportion Bought				Institutional Industry Proportion Bought			
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
$Intercept$	47.70	[33.16]	36.75	[29.16]	45.20	[58.65]	37.56	[46.34]
$LN\ (Size)$	-0.01	[-0.08]	0.00	[0.05]	0.61	[10.19]	0.69	[8.32]
$LN\ (BM)$	-2.64	[-6.85]	-2.24	[-5.25]	-0.43	[-3.51]	-0.37	[-2.54]
IND_RET_{t-1}	-6.92	[-1.68]	-6.24	[-1.01]	3.87	[3.35]	2.88	[2.65]
$IND\_RET_{t-3,t-2}$	-6.32	[-1.17]	1.34	[0.34]	2.41	[1.53]	2.20	[1.34]
$IND\_RET_{t-6,t-4}$	14.78	[3.77]	11.34	[2.32]	2.35	[1.28]	2.01	[1.38]
$IND\_RET_{t-12,t-7}$	44.19	[5.77]	40.23	[4.23]	3.26	[1.27]	3.02	[1.02]
$IND\_RET_{t-24,t-13}$	65.47	[7.83]	62.34	[5.32]	-5.38	[-1.44]	-4.33	[-1.02]
IND_PB_{t-1}			2.23	[3.92]			1.33	[2.54]
$IND\_PB_{t-3,t-2}$			1.99	[5.34]			0.79	[3.43]
$IND\_PB_{t-6,t-4}$			4.23	[4.28]			-0.36	[-0.73]
$IND\_PB_{t-12,t-7}$			2.03	[5.34]			0.29	[1.78]
$IND\_PB_{t-24,t-13}$			0.68	[2.49]			0.11	[0.23]
$Adjusted\ R^2$	0.26		0.35		0.20		0.28	

Panel B: Firm Proportion Bought								
	Retail Investor Firm Proportion Bought				Institutional Firm Proportion Bought			
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
$Intercept$	36.33	[46.05]	25.60	[20.48]	33.82	[53.55]	24.21	[22.24]
$LN\ (Size)$	0.76	[5.87]	1.13	[5.66]	1.29	[12.22	1.57	[16.09]
$LN\ (BM)$	-2.41	[-6.40]	-1.98	[-2.14]	0.01	[0.23]	0.00	[0.19]
RET_{t-1}	-8.61	[-12.31]	-5.74	[-3.10]	2.83	[5.17]	3.30	[8.35]
$RET_{t-3,t-2}$	-8.03	[-9.94]	-5.20	[-4.19]	3.58	[4.29]	3.91	[8.01]
$RET_{t-6,t-4}$	-0.46	[-0.70]	2.13	[2.85]	3.06	[3.34]	3.53	[6.21]
$RET_{t-12,t-7}$	11.65	[12.63]	10.79	[10.00]	1.69	[2.09]	1.79	[2.24]
$RET_{t-24,t-13}$	18.85	[15.83]	13.89	[9.97]	-3.73	[-3.85]	-2.74	[-3.13]
IND_RET_{t-1}	6.43	[3.42]	5.87	[2.27]	5.12	[4.06]	4.66	[2.69]
$IND\_RET_{t-3,t-2}$	4.72	[2.40]	1.27	[0.35]	-0.38	[-0.21]	2.23	[1.1]
$IND\_RET_{t-6,t-4}$	6.71	[2.77]	9.59	[2.48]	-3.38	[-1.45]	-4.18	[-2.19]
$IND\_RET_{t-12,t-7}$	24.85	[3.90]	23.69	[4.33]	-5.78	[-2.10]	-6.39	[-2.67]
$IND\_RET_{t-24,t-13}$	19.88	[4.20]	12.41	[2.39]	-21.14	[-5.61]	-23.01	[-4.89]
PB_{t-1}			9.30	[5.87]			3.56	[2.96]
$PB_{t-3,t-2}$			6.90	[3.37]			1.79	[2.65]
$PB_{t-6,t-4}$			2.41	[4.80]			2.05	[1.8]
$PB_{t-12,t-7}$			4.53	[3.37]			2.14	[2.73]
$PB_{t-24,t-13}$			6.62	[2.21]			1.59	[2.35]
$Adjusted\ R^2$	0.08		0.12		0.04		0.10	

Table 11
Returns on Portfolios Sorted on Past Industry Proportion Bought

This table sorts industries into portfolios based on the past n month industry proportion bought for both small and large trades. The industries most heavily sold (bought) over the prior n months are placed into portfolio 1 (5). We then examine the average monthly return on each portfolio over the subsequent n months. For each industry, we compute a value-weighted return. The portfolio returns is the average return across all the industries in the portfolio. Market-adjusted returns are the return on the portfolio less the value-weighted market return. Five-factor alphas are the intercept from a time-series regression where the dependent variable is the monthly return on the portfolio less the risk-free rate and the independent variables are the market, size, book-to-market, momentum, and industry momentum factors. The differences in return between quintile 5 and 1 is also reported, along with t-statistics in parentheses. The formation and holding period is 3 months in Panel A, 6 months in Panel B, and 12 months in Panel C.

Panel A: Three Months – Three Months

	Market-Adjusted Returns (%)			Five-Factor Alphas (%)		
Portfolio	Retail Traders	Institutions	Difference	Retail Traders	Institutions	Difference
1 (sold)	0.375	-0.056	0.431	0.324	-0.017	0.341
2	0.282	0.249	0.033	0.162	0.148	0.014
3	0.102	0.189	-0.087	0.210	0.269	-0.059
4	0.023	0.179	-0.156	0.002	0.139	-0.137
5 (bought)	-0.101	0.107	-0.208	-0.090	0.093	-0.183
B-S (5-1)	-0.476	0.163	-0.639	-0.414	0.110	-0.524
	(-3.46)	(1.30)	(-4.36)	(-3.19)	(1.03)	(-3.79)

Panel B: Six Months – Six Months

	Market-Adjusted Returns (%)			Five-Factor Alphas (%)		
Portfolio	Retail Traders	Institutions	Difference	Retail Traders	Institutions	Difference
1 (sold)	0.249	-0.021	0.270	0.245	-0.005	0.260
2	0.299	0.290	0.009	0.202	0.125	0.077
3	0.190	0.186	0.004	0.174	0.302	-0.128
4	0.100	0.193	-0.093	0.118	0.175	-0.057
5 (bought)	-0.162	0.045	-0.207	-0.142	0.015	-0.157
B-S (5-1)	-0.411	0.066	-0.477	-0.387	0.020	-0.407
	(-2.59)	(0.59)	(-2.49)	(-2.31)	(0.18)	(-2.47)

Panel C: One Year – One Year

	Market-Adjusted Returns (%)			Five-Factor Alphas (%)		
Portfolio	Retail Traders	Institutions	Difference	Retail Traders	Institutions	Difference
1 (sold)	0.342	-0.032	0.374	0.274	-0.020	0.314
2	0.103	0.321	-0.218	0.135	0.359	-0.224
3	0.217	0.219	-0.002	0.263	0.242	0.021
4	0.066	0.142	-0.076	0.102	0.152	-0.050
5 (bought)	-0.043	0.082	-0.125	-0.070	-0.010	-0.060
B-S (5-1)	-0.385	0.114	-0.499	-0.344	0.010	-0.354
	(-2.22)	(0.97)	(-2.75)	(-2.12)	(0.45)	(-2.00)

Table 12
Returns on Portfolios Sorted on Past Week Industry Proportion Bought

Each week from January 4, 1983 to December 27th 2000, portfolios are formed on the basis of prior week retail investor (institutional) industry proportion bought. Panel A reports the returns on the portfolio during the ranking period. On each day of the ranking period, the value-weighted return for each industry is computed. The portfolio return is the average of the industry returns in the portfolio. Daily returns are compounded to yield a monthly return series. Market-adjusted returns are the difference between the portfolio return and the value-weighted market return. Five-factor alphas are the intercept from a time-series regression where the dependent variable is the monthly return on the portfolio less the risk-free rate and the independent variables are the market, size, book-to-market, momentum, and industry momentum factors. Panel B reports the returns on each portfolio over the subsequent one week. The returns for each portfolio are computed analogously. The differences in return between quintile 5 and 1 is also reported, along with t-statistics in parentheses.

Panel A: Contemporaneous Returns

	Market-Adjusted Returns (%)			Five-Factor Alphas (%)		
Portfolio	Retail Traders	Institutions	Difference	Retail Traders	Institutions	Difference
1 (sold)	-1.991	-3.543	1.552	-1.830	-3.279	1.449
2	-1.302	-2.320	1.018	-1.434	-2.283	0.849
3	0.021	1.203	-1.182	-0.109	1.403	-1.512
4	1.219	2.932	-1.173	1.432	3.232	-1.800
5 (bought)	2.721	5.439	-2.718	2.630	5.299	-2.669
B-S (5-1)	4.712	8.932	-4.270	4.460	8.578	-4.118
	‘(27.38)	‘(45.83)	-(23.39)	(24.79)	(42.43)	(-20.74)

Panel B: Subsequent Returns

	Market-Adjusted Returns (%)			Five-Factor Alphas (%)		
Portfolio	Retail Traders	Institutions	Difference	Retail Traders	Institutions	Difference
1 (sold)	-0.286	0.423	-0.709	-0.147	0.328	-0.475
2	-0.181	0.389	-0.570	-0.253	0.349	-0.602
3	0.323	-0.29	0.613	0.195	-0.092	0.287
4	0.121	-0.212	0.333	0.102	-0.232	0.334
5 (bought)	0.438	-0.129	0.567	0.495	-0.121	0.616
B-S (5-1)	0.724	-0.552	1.276	0.642	-0.449	1.091
	(5.30)	(-5.04)	(6.89)	(4.29)	(-3.42)	(5.73)

Table 13 Retail Investor Industry and Stock Selection

This table decomposes the performance of retail investor trading into two components: industry selection and stock selection. Portfolios are formed on the basis of prior retail investor firm-level proportion bought. The return of the portfolio (total return) is the value-weighted average of the stocks return in that portfolio. The industry return is computed by substituting the return of the stock in the portfolio by the value-weighted return of the industry to which that stock returns. Stock return is defined as the difference between the total return and the industry return. Market-adjusted returns are the difference between the portfolio return and the value-weighted market return. Five-factor alphas are the intercept from a time-series regression where the dependent variable is the monthly return on the portfolio less the risk-free rate and the independent variables are the market, size, book-to-market, momentum, and industry momentum factors. The differences in return between quintile 5 and 1 is also reported, along with t-statistics in parentheses. The formation and holding period is 1 week in Panel A, 3 months in Panel B, 6 months in Panel C, and 12 months in Panel D.

Panel A: One Week – One Week						
	Market-Adjusted Returns (%)			Five-Factor Alphas (%)		
Portfolio	Total Ret.	Industry Ret.	Stock Ret.	Total Ret.	Industry Ret.	Stock Ret.
1 (sold)	-0.326	-0.192	-0.134	-0.289	-0.153	-0.136
2	-0.239	-0.089	-0.150	-0.293	-0.131	-0.162
3	0.149	-0.054	0.203	0.123	-0.081	0.204
4	0.189	-0.015	0.204	0.201	0.012	0.189
5 (bought)	0.480	0.262	0.218	0.502	0.272	0.230
B-S (5-1)	0.806	0.454	0.352	0.791	0.425	0.366
	(6.49)	(3.47)	(2.39)	(6.21)	(3.18)	(2.76)
Panel B: Three Months – Three Months						
	Market-Adjusted Returns (%)			Five-Factor Alphas (%)		
Portfolio	Total Ret.	Industry Ret.	Stock Ret.	Total Ret.	Industry Ret.	Stock Ret.
1 (sold)	0.426	0.301	0.125	0.440	0.323	0.117
2	0.229	0.153	0.076	0.203	0.129	0.074
3	0.119	0.029	0.090	0.020	-0.021	0.041
4	-0.062	-0.025	-0.037	-0.123	-0.012	-0.111
5 (bought)	-0.103	-0.006	-0.097	-0.150	0.021	-0.171
B-S (5-1)	-0.529	-0.307	-0.222	-0.590	-0.302	-0.288
	(-3.45)	(-2.12)	(-1.53)	(-3.89)	(-2.03)	(-1.73)
Panel C: Six Months – Six Months						
	Market-Adjusted Returns (%)			Five-Factor Alphas (%)		
Portfolio	Total Ret.	Industry Ret.	Stock Ret.	Total Ret.	Industry Ret.	Stock Ret.
1 (sold)	0.486	0.302	0.184	0.502	0.248	0.254
2	0.186	0.199	-0.013	0.142	0.172	-0.030
3	-0.035	-0.002	-0.033	-0.065	0.029	-0.094
4	0.019	-0.015	0.034	0.002	0.071	-0.069
5 (bought)	-0.013	0.005	-0.018	-0.034	-0.091	0.057
B-S (5-1)	-0.499	-0.297	-0.202	-0.536	-0.339	-0.197
	(-2.98)	(-2.00)	(-1.47)	(-3.19)	(-2.10)	(-1.25)
Panel D: One Year – One Year						
	Market-Adjusted Returns (%)			Five-Factor Alphas (%)		
Portfolio	Total Ret.	Industry Ret.	Stock Ret.	Total Ret.	Industry Ret.	Stock Ret.
1 (sold)	0.402	0.298	0.104	0.391	0.285	0.106
2	0.135	0.177	-0.042	0.119	0.159	-0.040
3	0.089	-0.002	0.091	-0.020	-0.021	0.001
4	-0.008	-0.015	0.007	-0.020	0.012	-0.032
5 (bought)	0.009	0.050	-0.041	-0.015	0.021	-0.002
B-S (5-1)	-0.393	-0.248	-0.145	-0.405	-0.264	-0.141
	(-2.09)	(-1.76)	(-1.07)	(-2.31)	(-1.81)	(-0.97)

Table 14
Summary sample statistics

The table shows fund portfolio statistics over select years during the 1980-2006 sample period. We define industry using two-digit SIC codes.

Year	Number of funds	Median assets ($M)	Median stocks	Median industries
1980	382	38	33	17
1985	464	84	39	18
1990	604	87	41	21
1995	2,002	98	49	22
2000	2,083	245	58	21
2006	1,478	431	69	28
1980-2006	3,562	155	52	22

Table 15
Factor model estimates

We estimate single- and multi-factor model regressions over quarterly horizons:

$$r_{p,t} = \alpha_p + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t}\,, \tag{1}$$

$$r_{pi,t} = \alpha_{pi} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t}\,, \tag{2}$$

$$r_{p,t} = \alpha_{ps} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \beta_{pi} r^{o}_{pi,t} + \varepsilon_{p,t}\,, \tag{3}$$

where $r_{p,t}$ represents fund gross excess returns, $r_{pi,t}$ represents industry excess returns, $r_{j,t}$ represents market, size, value, or momentum factors, and $r^{o}_{pi,t}$ represents an industry factor. The table reports the mean of the time series of mean quarterly alphas. *t*-statistics for the estimates are shown in parenthesis. Alphas are reported as daily percentages. The sample consists of 3,562 funds over a 1980-2006 sample period.

Skill type	α	β_m	β_{smb}	β_{hml}	β_{umd}	β_i	R^2
			Panel A. Single-factor				
α_p	0.0048 (2.28)	1.009					0.712
α_{pi}	0.0010 (1.33)	1.015					0.816
α_{ps}	0.0037 (1.71)	1.008				1.033	0.890
			Panel B. Three-factor				
α_p	0.0047 (2.12)	1.003	0.247	-0.019			0.775
α_{pi}	0.0011 (1.39)	1.019	0.118	-0.002			0.831
α_{ps}	0.0034 (1.58)	1.002	0.249	-0.018		1.022	0.914
			Panel C. Four-factor				
α_p	0.0049 (2.23)	1.004	0.222	-0.009	0.025		0.789
α_{pi}	0.0012 (1.38)	1.016	0.106	-0.005	0.012		0.852
α_{ps}	0.0035 (1.67)	1.005	0.221	-0.009	0.024	1.048	0.919

Table 16 Performance persistence

The table shows average daily percentage performance estimates during various quarters of a three-year post-ranking horizon for deciles of funds sorted according to fund performance estimated over the preceding quarter. Total alpha (Panel A) is the intercept, α_p, in a standard regression model:

$$r_{p,t} = \alpha_p + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t} \,. \tag{1}$$

Industry alpha (Panel B) is the intercept, α_{pi}, in a regression model where we use industry returns, $r_{pi,t}$, as the regressand:

$$r_{pi,t} = \alpha_{pi} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t} \,. \tag{2}$$

Stock alpha (Panel C) is the intercept, α_{ps}, in a regression model that includes an additional industry factor, $r^{o}_{pi,t}$:

$$r_{p,t} = \alpha_{ps} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \beta_{pi} r^{o}_{pi,t} + \varepsilon_{p,t} \,. \tag{3}$$

$r_{p,t}$ represents fund gross excess returns, $r_{j,t}$ represents market, size, value, or momentum factors, and $r^{o}_{pi,t}$ represents an industry factor. The table reports average daily percentage returns associated with each post-ranking quarter, rather than the cumulative returns through each quarter. All results are based on the four-factor model. "10" refers to the best past performance decile, and "1" refers to the worst past performance decile. *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively. The sample consists of 3,562 funds over a 1980-2006 sample period.

Ranking	Post-ranking quarter							
Decile	1	2	3	4	6	8	10	12
				Panel A. Total alpha				
10	0.0114	0.0120	0.0132	0.0122	0.0112	0.0103	0.0090	0.0091
9	0.0087	0.0091	0.0082	0.0082	0.0078	0.0074	0.0069	0.0066
8	0.0081	0.0079	0.0061	0.0056	0.0058	0.0056	0.0053	0.0052
7	0.0064	0.0062	0.0047	0.0052	0.0053	0.0051	0.0046	0.0049
6	0.0063	0.0058	0.0041	0.0041	0.0044	0.0045	0.0044	0.0046
5	0.0040	0.0043	0.0039	0.0041	0.0046	0.0047	0.0044	0.0044
4	0.0047	0.0036	0.0029	0.0033	0.0038	0.0037	0.0040	0.0043
3	0.0018	0.0023	0.0030	0.0034	0.0045	0.0046	0.0044	0.0047
2	-0.0006	-0.0004	0.0021	0.0024	0.0032	0.0038	0.0042	0.0046
1	-0.0012	0.0001	0.0025	0.0028	0.0044	0.0045	0.0053	0.0058
10-1	0.0126***	0.0119***	0.0107***	0.0094***	0.0068***	0.0058**	0.0037	0.0033
Spearman	0.964***	0.988***	0.976***	0.964***	0.891***	0.806***	0.576*	0.479

Table 3 continued.

Ranking Decile	Post-ranking quarter 1	2	3	4	6	8	10	12
				Panel B. Industry alpha				
10	0.0045	0.0046	0.0042	0.0038	0.0030	0.0025	0.0023	0.0021
9	0.0030	0.0031	0.0030	0.0029	0.0026	0.0021	0.0021	0.0024
8	0.0030	0.0030	0.0025	0.0026	0.0023	0.0023	0.0026	0.0026
7	0.0018	0.0016	0.0022	0.0022	0.0022	0.0019	0.0020	0.0023
6	0.0015	0.0021	0.0015	0.0018	0.0020	0.0017	0.0017	0.0020
5	0.0008	0.0008	0.0009	0.0013	0.0015	0.0013	0.0014	0.0015
4	0.0007	0.0007	0.0008	0.0012	0.0015	0.0012	0.0013	0.0015
3	0.0009	0.0004	0.0005	0.0006	0.0011	0.0010	0.0010	0.0012
2	-0.0007	-0.0003	-0.0000	0.0000	0.0006	0.0004	0.0008	0.0008
1	-0.0036	-0.0034	-0.0021	-0.0020	-0.0006	-0.0007	-0.0003	-0.0001
10-1	0.0081***	0.0080***	0.0063**	0.0058**	0.0036	0.0032	0.0026	0.0022
Spearman	0.964***	0.988***	1.000***	1.000***	1.000***	0.988***	0.952***	0.903***
				Panel C. Stock alpha				
10	0.0075	0.0078	0.0081	0.0072	0.0069	0.0063	0.0066	0.0056
9	0.0055	0.0057	0.0042	0.0043	0.0051	0.0051	0.0049	0.0042
8	0.0063	0.0051	0.0038	0.0036	0.0033	0.0032	0.0032	0.0034
7	0.0041	0.0056	0.0031	0.0031	0.0031	0.0029	0.0016	0.0021
6	0.0015	0.0024	0.0041	0.0032	0.0023	0.0025	0.0026	0.0029
5	0.0035	0.0037	0.0025	0.0027	0.0031	0.0037	0.0035	0.0030
4	0.0038	0.0021	0.0020	0.0027	0.0027	0.0022	0.0021	0.0029
3	0.0012	0.0012	0.0026	0.0034	0.0032	0.0038	0.0039	0.0033
2	0.0002	-0.0002	0.0015	0.0026	0.0029	0.0031	0.0036	0.0041
1	0.0014	0.0022	0.0032	0.0032	0.0041	0.0043	0.0045	0.0047
10-1	0.0061**	0.0056**	0.0049*	0.0040	0.0028	0.0020	0.0021	0.0009
Spearman	0.875***	0.842***	0.733**	0.665**	0.406	0.273	0.115	0.091

Table 17
Predicting total alpha with total alpha, industry alpha, or stock alpha

The table shows results from cross-sectional regressions of total alpha vs. past performance:

$$\alpha_{p,t} = a + b\alpha_{p,t-1} + \varepsilon_{p,t}, \quad (4)$$

$$\alpha_{p,t} = a + b\alpha_{pi,t-1} + \varepsilon_{p,t}, \quad (5)$$

$$\alpha_{p,t} = a + b\alpha_{ps,t-1} + \varepsilon_{p,t}, \quad (6)$$

and

$$\alpha_{p,t} = a + b\alpha_{pi,t-1} + c\alpha_{ps,t-1} + \varepsilon_{p,t}. \quad (7)$$

where α_p, α_{pi}, and α_{ps} refer to total alpha, industry alpha, and stock alpha, respectively, estimated over quarterly horizons:

$$r_{p,t} = \alpha_p + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t}, \quad (1)$$

$$r_{pi,t} = \alpha_{pi} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t}, \quad (2)$$

$$r_{p,t} = \alpha_{ps} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \beta_{pi} r^{o}_{pi,t} + \varepsilon_{p,t}. \quad (3)$$

$r_{p,t}$ represents fund gross excess returns, $r_{pi,t}$ represents industry excess returns, $r_{j,t}$ represents market, size, value, or momentum factors, and $r^{o}_{pi,t}$ represents an industry factor. The table reports mean coefficient estimates, Fama MacBeth *t*-statistics (in parenthesis), and mean r-squares. The sample consists of 3,562 funds over a 1980-2006 sample period.

	Single-factor				Three-factor				Four-factor			
a	0.004	0.005	0.005	0.004	0.006	0.006	0.006	0.006	0.007	0.006	0.007	0.006
	(1.57)	(1.73)	(1.60)	(1.63)	(2.90)	(2.97)	(3.02)	(2.88)	(3.40)	(3.43)	(3.46)	(5.41)
α_p	0.045				0.058				0.030			
	(2.30)				(3.71)				(2.32)			
α_{pi}		0.083		0.080		0.105		0.102		0.053		0.051
		(2.25)		(2.20)		(3.57)		(3.56)		(2.10)		(2.12)
α_{ps}			0.029	0.027			0.035	0.034			0.012	0.014
			(1.43)	(1.41)			(2.28)	(2.11)			(0.91)	(0.96)
R^2	0.043	0.040	0.023	0.059	0.031	0.040	0.016	0.042	0.017	0.020	0.011	0.032

Table 18
Future total alpha of deciles sorted according to industry or stock alpha

The table shows average daily percentage total alpha estimates during various quarters of a three-year post-ranking horizon for deciles of funds sorted according to industry alpha (Panel A) or stock alpha (Panel B) estimated over the preceding quarter. Total alpha is the intercept, α_p, in a standard regression model:

$$r_{p,t} = \alpha_p + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t} \,. \tag{1}$$

Industry alpha is the intercept, α_{pi}, in a regression model where we use industry returns, $r_{pi,t}$, as the regressand:

$$r_{pi,t} = \alpha_{pi} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t} \,. \tag{2}$$

Stock alpha is the intercept, α_{ps}, in a regression model that includes an additional industry factor, $r^{o}_{pi,t}$:

$$r_{p,t} = \alpha_{ps} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \beta_{pi} r^{o}_{pi,t} + \varepsilon_{p,t} \,. \tag{3}$$

$r_{p,t}$ represents fund gross excess returns, $r_{j,t}$ represents market, size, value, or momentum factors, and $r^{o}_{pi,t}$ represents an industry factor. The table reports average daily percentage returns associated with each post-ranking quarter, rather than the cumulative returns through each quarter. All results are based on the four-factor model. "10" refers to the best past performance decile, and "1" refers to the worst past performance decile. *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively. The sample consists of 3,562 funds over a 1980-2006 sample period.

	Post-ranking quarter					
Decile	1	2	3	4	8	12
	Panel A. Industry alpha sort					
10	0.0137	0.0131	0.0135	0.0125	0.0108	0.0099
9	0.0123	0.0115	0.0097	0.0086	0.0082	0.0078
2	-0.0008	-0.0001	0.0011	0.0022	0.0025	0.0044
1	-0.0010	-0.0003	0.0005	0.0020	0.0036	0.0052
10-1	0.0146***	0.0134***	0.0130***	0.0105***	0.0072**	0.0047
	Panel B. Stock alpha sort					
10	0.0109	0.0107	0.0103	0.0098	0.0086	0.0078
9	0.0076	0.0071	0.0073	0.0079	0.0065	0.0075
2	-0.0003	0.0017	0.0019	0.0034	0.0038	0.0039
1	0.0011	0.0025	0.0027	0.0039	0.0042	0.0046
10-1	0.0098***	0.0082***	0.0076**	0.0059*	0.0044	0.0032

Table 19
Performance and fund portfolio size

The table shows average daily percentage performance estimates over a quarterly horizon for deciles of funds sorted according to fund portfolio size at the end of the previous quarter. Total alpha (Panel A) is the intercept, α_p, in a standard regression model:

$$r_{p,t} = \alpha_p + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t} \; . \tag{1}$$

Industry alpha (Panel B) is the intercept, α_{pi}, in a regression model where we use industry returns, $r_{pi,t}$, as the regressand:

$$r_{pi,t} = \alpha_{pi} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \varepsilon_{p,t} \; . \tag{2}$$

Stock alpha (Panel C) is the intercept, α_{ps}, in a regression model that includes an additional industry factor, $r^o_{pi,t}$:

$$r_{p,t} = \alpha_{ps} + \sum_{j=1}^{k} \beta_{pj} r_{j,t} + \beta_{pi} r^o_{pi,t} + \varepsilon_{p,t} \; . \tag{3}$$

$r_{p,t}$ represents fund gross excess returns, $r_{j,t}$ represents market, size, value, or momentum factors, and $r^o_{pi,t}$ represents an industry factor. “10” refers to the largest size decile, and “1” refers to the smallest size decile. *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively. The sample consists of 3,562 funds over a 1980-2006 sample period.

Size decile	Single-factor	Three-factor	Four-factor
	Panel A. Total alpha		
10	-0.0004	0.0032	0.0037
9	0.0023	0.0045	0.0046
8	0.0018	0.0040	0.0033
7	0.0038	0.0043	0.0043
6	0.0051	0.0055	0.0051
5	0.0064	0.0049	0.0055
4	0.0062	0.0042	0.0049
3	0.0071	0.0049	0.0055
2	0.0066	0.0045	0.0048
1	0.0081	0.0076	0.0078
10-1	-0.0085***	-0.0044*	-0.0041*
Spearman	-0.964***	-0.588*	-0.782***
	Panel B. Industry alpha		
10	-0.0002	0.0019	0.0022
9	0.0003	0.0016	0.0016
8	0.0005	0.0017	0.0015
7	0.0007	0.0012	0.0012
6	0.0013	0.0013	0.0014
5	0.0011	0.0010	0.0007
4	0.0014	0.0003	0.0009
3	0.0023	0.0010	0.0011
2	0.0020	0.0008	0.0008
1	0.0007	0.0000	0.0008
10-1	-0.0009	0.0019	0.0014
Spearman	-0.794***	0.915***	0.848***

Figure 1
Large-trade dollar volume cutoffs

This figure shows cross-sectional average large-trade dollar volume cutoffs over time.

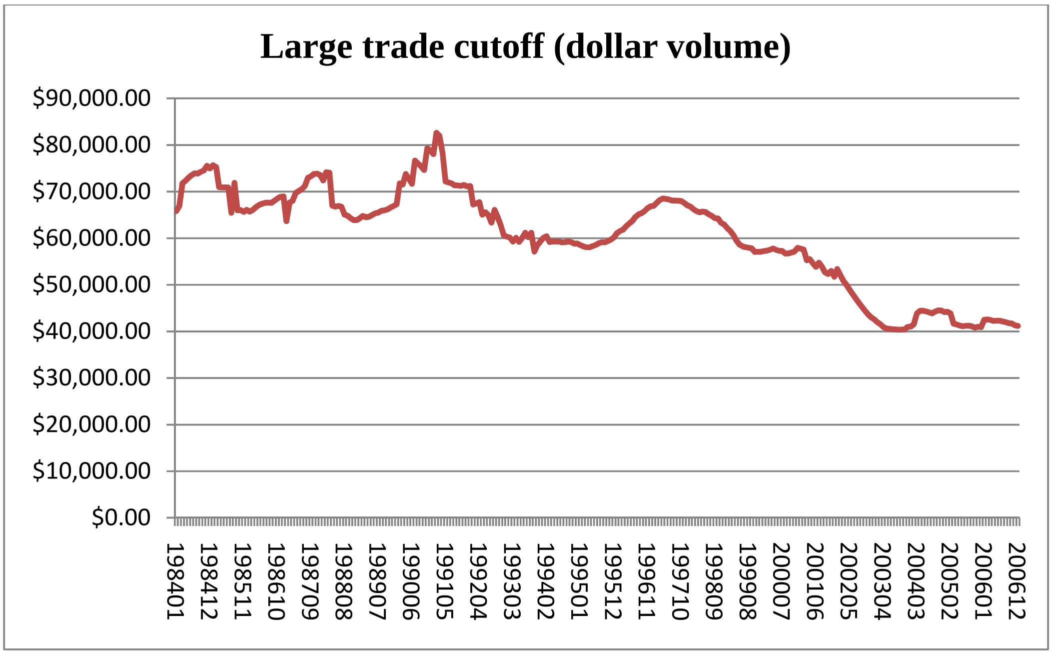


Figure 2

Cross Sectional Correlation of Industry Order Imbalance

Each week from January 4, 1983 to December 27, 2000 we compute retail investor (institutional) industry order imbalance. This figure reports the time series average of the cross sectional correlations between retail investor (institutional) industry order imbalance in week *t*, and week *t+x*. The x axis represents different horizons.

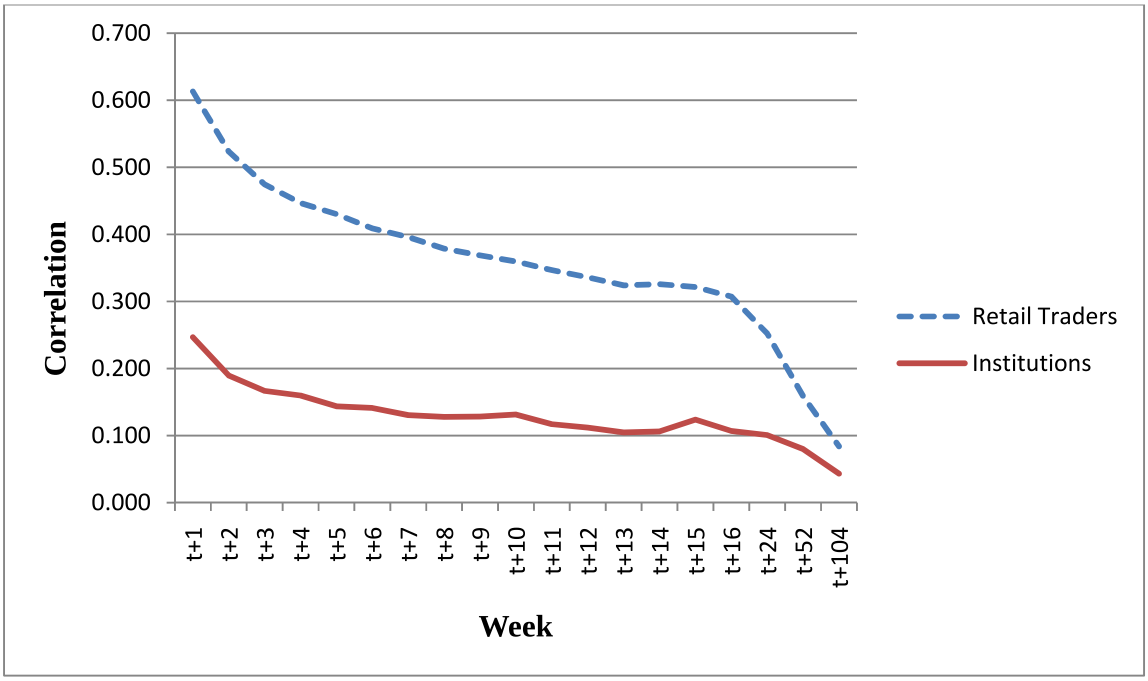


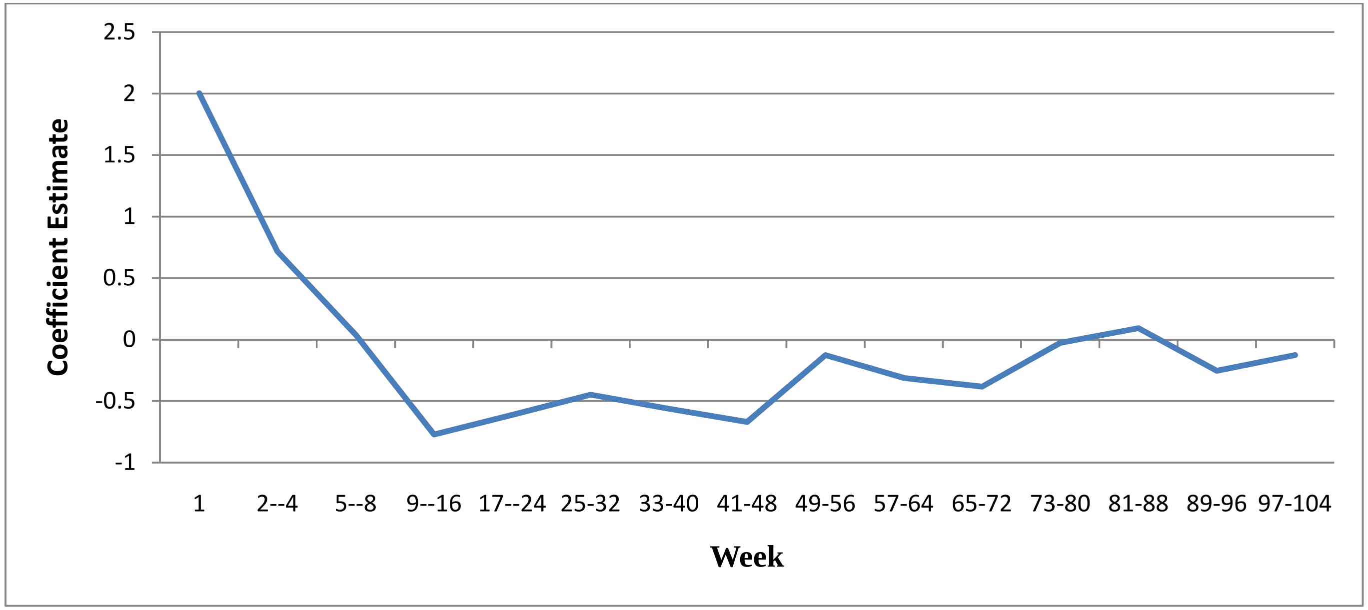
Figure 3
The Effect of Past Industry Proportion Bought on Industry Returns

Each week from January 4, 1984 to through December 27, 2000 we fun the following cross-sectional regression:

$$IndRet_{it} = a_o + b_1 Ind\_PB_{it-1} + b_2 Ind\_PB_{i,t-4,t-2} + b_3 Ind\_PB_{i,t-8,t-5} \sum_{w=9}^{97\ by\ 8} b_{t-w,t-w-7}\, Ind\_PB_{it-w,t-w-7} + c_1 MVE_{it} + d_1 BM_{it}$$
$$+ \sum_{w=1}^{4} e_{t-w,}\, Ind\_Ret_{it-w,} + f_1 Ind\_Ret_{it-52,t-5} + g_1 Ind\_Ret_{it-104,t-53} + \varepsilon_i$$

The dependent variable is the industry return in week *t.* The independent variables includes the industry proportion bought over the prior week, prior two to four weeks, prior 5 to 8 weeks, and subsequent 8 week periods, beginning with the prior 9 to 16 weeks and ending over the prior 97 to 104 weeks. Other control variables include the average industry market cap, the average industry book to market, and prior industry returns. The figure presents the mean coefficient estimates on the lagged industry proportion bought variables. The coefficients and t-statistics are based on the time-series mean and time-series standard deviation of the cross-sectional estimates. Panel A reports the results for retail investor industry proportion bought. Panel B reports the results for institutional industry proportion bought.

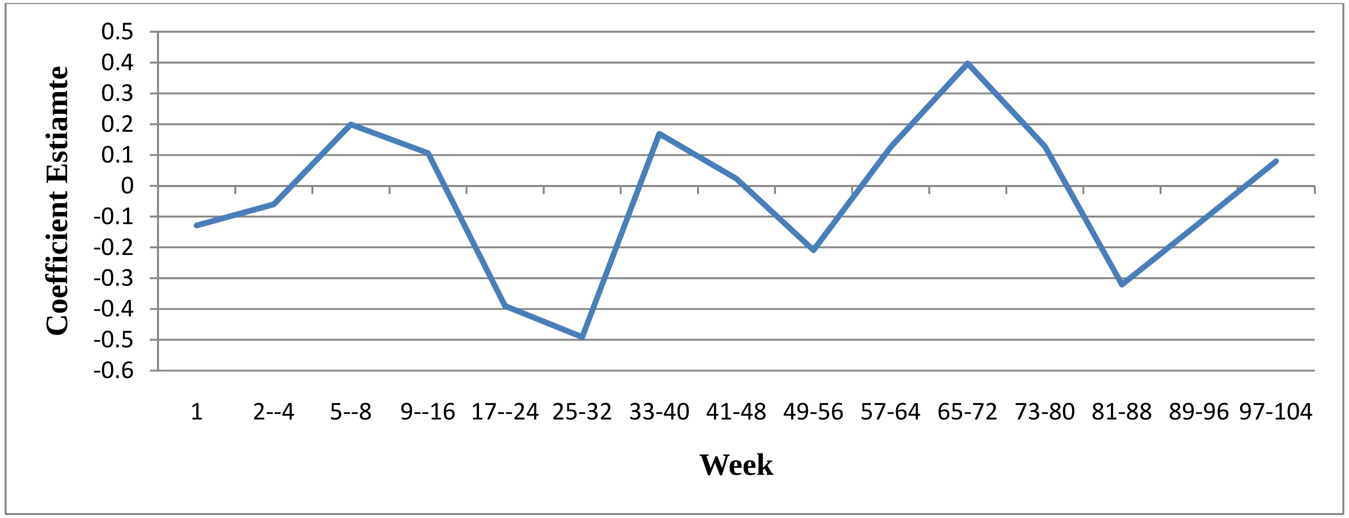
Panel A: Coefficient Estimates on Lagged Industry Proportion Bought of Retail Investors



	1	2--4	5--8	9--16	17--24	25-32	33-40	41-48	49-56	57-64	65-72	73-80	81-88	89-96	97-104
coef	2.002	0.716	0.038	-0.772	-0.612	-0.448	-0.562	-0.67	-0.126	-0.312	-0.383	-0.026	0.092	-0.254	-0.126
t-stat	12.32	3.36	0.68	-6.32	-4.12	-3.32	-4.42	-3.96	-1.36	-2.57	-2.02	-0.19	1.31	-2.38	-0.58

	$SIZE$	BM	Ind_Ret_{t-1}	Ind_Ret_{t-2}	Ind_Ret_{t-3}	Ind_Ret_{t-4}	$Ind\_Ret_{t-52,t-5}$	$Ind\_Ret_{t-104,t-53}$
coef	0.011	0.125	0.017	0.015	0.012	0.002	0.020	0.050
t-stat	1.39	3.15	2.95	2.32	2.48	1.07	3.39	1.59

Panel B: Coefficient Estimates on Lagged Industry Proportion Bought of Institutions



	1	2--4	5--8	9--16	17--24	25-32	33-40	41-48	49-56	57-64	65-72	73-80	81-88	89-96	97-104
coef	-0.129	-0.06	0.199	0.106	-0.39	-0.491	0.169	0.023	-0.209	0.126	0.397	0.128	-0.321	-0.121	0.08
t-stat	-2.16	-1.09	1.39	1.09	-2.75	-2.32	1.03	0.36	-1.15	0.78	2.49	1.68	-2.67	-1.10	0.98

	$SIZE$	BM	Ind_Ret_{t-1}	Ind_Ret_{t-2}	Ind_Ret_{t-3}	Ind_Ret_{t-4}	$Ind\_Ret_{t-52,t-5}$	$Ind\_Ret_{t-104,t-53}$
coef	0.015	0.116	0.015	0.012	0.010	0.005	0.022	0.008
t-stat	1.58	2.77	2.69	2.51	2.28	1.45	3.64	1.79