

# MASS PSYCHOLOGY AND RETURN COMOVEMENTS: THE CASE OF RETAIL TRADES

Alok Kumar

*University of Notre Dame*

*Mendoza College of Business*

Charles M.C. Lee

*Cornell University*

*Johnson Graduate School of Management*

March 22, 2004

(Comments are welcomed)

---

Please address all correspondence to Alok Kumar, 239 Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556; *Phone*: 574-631-0354; *email*: akumar@nd.edu OR Charles M.C. Lee, Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853; *Phone*: 607-255-6255; *email*: CL86@cornell.edu. We would like to thank Robert Battalio, Xiang Cai, Shane Corwin, Richard Frankel, William Goetzmann, Hans Heidle, Dong Hong, Zoran Ivkovich, George Korniotis, J. Spencer Martin, Robert Masson, Victor McGee, Ted O'Donoghue, Vicente Pons, Paul Schultz, Mark Seasholes, Kent Womack, and Jeff Wurgler, as well as seminar participants at MIT Sloan School, the 2003 WFA Meeting in Los Cabos and University of Notre Dame for helpful discussions and comments. We thank Itamar Simonson for making the investor data available to us and Terrance Odean for answering numerous questions about the investor database. All remaining errors and omissions are our own.

# MASS PSYCHOLOGY AND RETURN COMOVEMENTS: THE CASE OF RETAIL TRADES

## **Abstract**

This study documents the existence of a common component in the buy-sell activities of retail investors, and evaluates its impact on stock returns. Our analysis is based on more than 1.85 million buy and sell transactions made by over 60,000 retail investors in a six year period. We show that retail trades are systematically correlated - i.e., individual investors buy (or sell) stocks in concert with each other. Moreover, a factor based on this common directional behavior explains return comovements, particularly for stocks with high retail concentrations that are also costly to arbitrage. Collectively, our results support a role for mass psychology in returns formation.

**Keywords:** Retail sentiment; Correlated trading; Habitat-based comovement; Mass psychology.

**JEL Classification:** G12, G14.

## **I Introduction**

The tendency among market researchers, and in the financial press, is to think of the market as a single unit. This notion permeates our vocabulary and is well entrenched in our thinking. The market is said to have gone up today, or the market has responded negatively to a firm's latest earnings report. In academic circles, market efficiency is defined in terms of the speed with which price (an aggregate measure) impounds information, and analytical models often feature "representative" agents. In all these cases, the market is regarded, perhaps too casually, as a monolithic whole.

An alternative view is to regard the market as consisting of various clienteles or informational subgroups. These subgroups respond to different information stimuli and may respond differently to the same information stimulus. The reactions of different subgroups are sometimes re-enforcing and sometimes off-setting. To the extent that market subgroups respond to different information signals, or have different time-varying liquidity needs, their trading behavior will not be perfectly correlated. These differences, under constrained arbitrage conditions, can generate pricing effects that will appear anomalous in traditional asset pricing models.

In this study, we test a returns generating model in which retail investors represent a significant clientele or informational subgroup.<sup>1</sup> The U.S. equity market is characterized by wide-spread direct stock ownership by individual investors. Extant evidence shows that retail investors spend far less time on investment analysis, engage in attention-based trading, and typically rely on a different set of information sources from their professional counterparts.<sup>2</sup> If retail investors'

---

<sup>1</sup> This concept underpins much of the noise trader literature, as formally modeled by, for example, Delong, Shleifer, Summers, and Waldmann (DSSW; 1990) and Barberis, Shleifer, and Wurgler (BSW; 2003).

<sup>2</sup> NYSE's Share Ownership Survey and the Securities Industry Association Investor Activity Report contain information on individual stock ownership and trading frequency. Lease, Lewellen, and Schlarbaum (1974), Lewellen, Lease, and Schlarbaum (1977), Yunker and Krehbiel (1974), Shiller and Pound (1989), and Frieder and Subrahmanyam (2002) discuss individual investor decision styles and information sources. Odean (1999) and Barber and Odean (2000, 2001) show retail investors trade too often, particularly around news events. Finally, Lee (1992), Bhattacharya (2001), and Battalio and Mendenhall (2003) show that small traders' response to earnings news differ sharply from those of large traders.

buy-sell patterns do not move in lock-step with overall market movements, assets in market segments dominated by these investors could be characterized by apparent pricing anomalies associated with their trading activities.

We use the personal trading records of individual investors to explore the effect of retail trading on stock returns. Our analysis is based on a database of more than 1.85 million buy and sell transactions made by over 60,000 retail clients of a large discount brokerage firm over a six year period (1991-96). In the first part of the study, we examine whether the buy-sell activities of these retail investors contain a common directional component. In the second part of the study, we form a mimicking long-short portfolio based on the direction of these retail trades, and evaluate the impact of retail investor trading on comovement in stock returns.

Our results show that the trades of these investors are systematically correlated - i.e., individual investors tend to buy (or sell) stocks in concert with each other. Specifically, we document a strong positive correlation in their buy-sell imbalance (BSI) that is orthogonal to overall market movements. We observe this common directional trading pattern across groups of different individual investors, and also in non-overlapping portfolios of different stocks. In other words, when some investors are buying (selling) stocks, other individuals also tend to be buying (selling). Similarly, when one group of stocks is being bought (sold) by retail investors, other groups also tend to be bought (sold). Collectively, these findings point to the existence of a common directional component in the trades of retail investors.

In the second part of the paper, we examine whether this systematic component can help explain comovement in stock returns. Specifically, we construct a market-wide retail investor sentiment factor (ISEN) consisting of the difference between the returns on an equal-weighted portfolio of stocks with high BSI and the returns on a portfolio of stocks with low BSI. This retail investor sentiment factor (ISEN) is a “change-in-sentiment” indicator designed to capture monthly shifts

in retail sentiment across different baskets of stocks.<sup>3</sup> To ensure that ISEN in a given month can be constructed *ex ante*, we use the BSI from the *previous* month to compute the current month ISEN.

Our results show that ISEN has significant incremental ability to explain comovement in returns, even after controlling for the market excess return (RMRF), the size factor (SMB), the book-to-market factor (HML), and the momentum factor (UMD). The sensitivity of firms to ISEN is most pronounced among stocks with a high retail concentration, or stocks in which retail trading activities are most intense. Specifically, we find that small firms, lower priced firms, firms with low institutional ownership, and value firms, all have positive loadings on ISEN.

Collectively, these findings are relevant to the debate on whether mass psychology plays a role in financial markets. The traditional case against a role for mass psychology in markets is based on two key assertions: (1) the cognitive foibles committed by individuals do not aggregate across the investing populous (i.e., individual irrationalities do not result in systematic directional behavior across large groups of investors); and (2) even if systematic noise trading exists, an army of rational arbitrageurs stands ready to offset this behavior, leaving prices unaffected.<sup>4</sup>

Despite their importance to the study of markets, these assertions are surprisingly difficult to test empirically. First, to evaluate whether individual behavior aggregate, we need to secure personal trading records for a large sample of investors. Second, to assert a role for investor sentiment, we need to demonstrate that these investors trade in concert. Third, if we uncover a common component in their trading behavior, we still need to show that their collective action has pricing implications - i.e., arbitrage forces do not fully nullify their collective behavior.

---

<sup>3</sup> Note that in contrast to our “change-in-sentiment” measure, Baker and Wurgler (2003) provide a “level-of-sentiment” measure designed to capture the aggregate level of investor sentiment across the entire market.

<sup>4</sup> See Shiller (1984), Shleifer (2000), or Lee (2001) for a more complete treatment.

Our analysis speaks directly to these issues, and our results suggest that, at least in the case of retail investors, both assertions are questionable. Specifically, we find that retail trades do aggregate across individuals, and that the collective action of these individuals can influence stock returns. Our analysis does not allow us to identify the precise driver of retail investor trading - i.e., whether their time-varying preference is due to liquidity concerns, risk aversion, or irrational sentiment. However, our results are broadly suggestive of a role for mass psychology in the study of financial markets.

Our results also provide support for friction- or sentiment-based theories of returns comovement, such as Barberis, Shleifer, and Wurgler (BSW; 2003). BSW presents a clientele-based explanation for the return patterns observed around the inclusion or deletion of a stock from the S&P500 stock index. Several aspects of our result are supportive of their theory. First, unlike traditional risk factors, ISEN is derived from the trading activities of a specific subgroup of the investing populous. We find that ISEN is not highly correlated with either Fama-French (1993) risk factors such as SMB, HML, and UMD, or macro-economic variables associated changes in market fundamentals (Chen, Roll and Ross (1986)). Moreover, consistent with the “habitat” version of the BSW model, we find that stocks preferred by retail investors are the ones most sensitive to the ISEN factor. Specifically, we show that retail investors concentrate in small, low-priced, high B/M, and low institutionally-owned stocks, and that these are the stocks with significant ISEN factor loadings.

We also examine a further prediction of their model. Specifically, BSW argue that stocks with the steepest short-run demand curves will be most difficult to arbitrage, and will be most affected by changes in investor sentiment. Therefore, even among firms that make up the natural “habitat” of retail investors (i.e., small firms, low-priced firms, value firms, and low institutionally-owned firms), the sensitivity of returns to ISEN should be a function of arbitrage

costs. Using the idiosyncratic risk of a stock as a measure of arbitrage cost,<sup>5</sup> we show that the factor loadings on ISEN are a (nearly monotonic) function of arbitrage costs. Across all four measures of retail investor habitat (i.e., small, low-priced, high B/M, and low institutionally-owned stocks), the firms that are more difficult to arbitrage exhibit much higher ISEN loadings.

Finally, our study is related to the literature on investor sentiment and closed-end fund discounts. Delong, Shleifer, Summers, and Waldmann (DSSW; 1990) conjecture that, because shares of closed-end funds are held primarily by individual investors, the discounts on these funds capture the differential sentiment of retail investors. Consistent with this view, Lee, Shleifer, and Thaler (LST; 1991) find that the returns of stocks with lower institutional ownership and lower market capitalization are positively correlated with changes in closed-end fund discounts. The LST finding proved controversial and has spawned a number of follow-up studies.<sup>6</sup> A key point of contention is whether closed-end fund discounts actually capture individual investor behavior. We shed light on this debate by providing direct evidence linking retail trades and stock returns. Consistent with LST, our results show that the directional trades of individual investors are associated with the excess returns of small, low-institutionally owned, firms, particularly if these firms are also difficult to arbitrage.

The remainder of our paper is organized as follows. In Section II, we develop a clientele-based returns model in more detail and discuss links to related literature. In Section III, we discuss model implementation, research methodology, and data issues. In Sections IV, V, and VI, we test the model using retail trade data and stock returns. Finally, Section VII concludes with a general summary and discussion of the implications of our findings.

---

<sup>5</sup> See Wurgler and Zhuravskaya (2002) for a detailed discussion of this variable, and why it is a good proxy for cross-sectional differences in arbitrage costs. Like their study, we define idiosyncratic risk as the variance of the residual from the CAPM regression.

<sup>6</sup> See Chen, Kan and Miller (1993a, 1993b), Chopra, Lee, Shleifer, and Thaler (1993a, 1993b), Elton, Gruber, and Busse (1998), and Gemmill and Thomas (2002).

## II Hypothesis Development

The traditional view of returns comovement relies on the present value model of security valuation, where the current price of a stock reflects the present discounted value of a stream of future earnings. According to this view, the returns of two assets are correlated because changes in the assets' fundamental values are correlated. Demand shocks or shifts in investor sentiment have no role to play in this traditional view because such shocks are readily offset by the actions of arbitrageurs.

An alternative theory of returns comovement posits that stock prices are established through a dynamic interplay between noise traders and rational arbitrageurs (e.g., Shiller (1984), Shleifer and Summers (1999)). According to this view, in addition to innovations in fundamentals, factors such as the correlated trading activities of noise traders, can also induce comovement. Arbitrage forces may not fully absorb these correlated demand shocks due to a variety of factors.<sup>7</sup>

In this study, we test a particular form of the noise trader model in which individual (or retail) investor sentiment can affect stock returns. We have in mind a clientele-based model that closely parallels those presented in Bodurtha, Kim and Lee (BKL, 1995) and Barberis, Shleifer, and Wurgler (BSW; 2003). In these models, different investor groups have different natural “habitats,” or groups of stocks, that they restrict themselves to trading. Thus, the returns of individual stocks will reflect not only fundamental risk, but also the time-varying preferences (or “sentiment”) of important investor groups.<sup>8</sup>

---

<sup>7</sup> See Shleifer and Vishny (1997) for a theoretical exposition of this argument and Rashes (2001) for an example where arbitrage fails to discipline the market even in a very simple setting.

<sup>8</sup> For example, in BKL, some U.S. investors trade only domestic securities. As a result, the stock prices of closed-end country funds traded in the U.S. are affected not only by their Net Asset Value (NAV), but also by U.S. market movements. Similarly, BSW considers several variations of a returns comovement model. In the case of their “habitat” model, some investors only trade a specific subset of the available securities (i.e., stocks in their preferred “habitat”). As a result, stocks preferred by a given clientele exhibit comovement beyond that attributable to fundamental news.

Notationally, consider an economy that contains one riskless and  $2n$  risky assets. The riskless asset has perfectly elastic supply and the risky assets are in fixed supply. Risky asset  $i$  is a claim to a single liquidating dividend,  $D_{i,T}$ , to be paid at some future time  $T$ . This eventual dividend equals

$$D_{i,T} = D_{i,0} + \varepsilon_{i,1} + \dots + \varepsilon_{i,T}, \quad (1)$$

where  $D_{i,0}$  and  $\varepsilon_{i,t}$  are announced at time 0 and time  $t$ , respectively, and where

$$\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{2n,t})' \sim N(0, \Sigma_D), \text{ i.i.d. over time.}$$

The price of a share of risky asset  $i$  at time  $t$  is  $P_{i,t}$  and the return (for simplicity, we refer to the asset's change in price as its return) on the asset between time  $t-1$  and time  $t$  is

$$\Delta P_{i,t} \equiv P_{i,t} - P_{i,t-1}. \quad (2)$$

Suppose that some investors have preferred “habitats,” such that they concentrate their trades in a subset of the  $2n$  risky assets. In the context of this study, consider the case where assets  $1$  through  $n$  (category  $X$  stocks) are stocks that preferred by retail investors; and assets  $n+1$  through  $2n$  (category  $Y$  stocks) are those stocks preferred by institutional investors.<sup>9</sup>

Suppose now that in addition to fundamental news, both categories are affected by noise traders, who channel funds in and out of the categories depending on their time-varying preferences (i.e., sentiment) or liquidity needs. A simple representation for asset returns is then

---

<sup>9</sup> For parsimony, we assume each stock is affected *either* by retail or institutional sentiment (i.e., the “habitat” of the two subgroups are non-overlapping). Without loss of generality, we could alternatively assume that each stock is affected by *both* retail and institutional sentiment. In that case, the analysis carries through so long as retail and institutional investors cluster in sufficient concentrations within their respective habitats.

$$\Delta P_{i,t} = \varepsilon_{i,t} + \Delta u_{X,t}, \quad i \in X \quad (3)$$

$$\Delta P_{j,t} = \varepsilon_{j,t} + \Delta u_{Y,t}, \quad j \in Y \quad (4)$$

where

$$\begin{pmatrix} u_{X,t} \\ u_{Y,t} \end{pmatrix} \sim \mathbf{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_u^2 \begin{pmatrix} 1 & \rho_u \\ \rho_u & 1 \end{pmatrix} \right), \quad i.i.d. \text{ over time.} \quad (5)$$

Here,  $u_{X,t}$  can be thought of as a variable tracking the risk aversion, sentiment, or liquidity needs of retail investors, who concentrate their investing in the securities in habitat  $X$ . The returns of assets in habitat  $X$  are affected not only by news about cash flows ( $\varepsilon_{i,t}$ ) but also by changes in the risk aversion, say, of retail investors ( $\Delta u_{X,t}$ ). In the same spirit, returns of assets in habitat  $Y$  are affected not only by news about cash flows ( $\varepsilon_{j,t}$ ), but also by changes in the sentiment or risk aversion of institutional investors ( $\Delta u_{Y,t}$ ).

Note that this model does not require each investor to be either retail or institutional – i.e., we do not assume the counter-party to a trade initiated by a retail (institutional) investor is always an institutional (retail) investor.<sup>10</sup> Other investors who respond only to fundamental news can trade against either pure retail or pure institutional investors. The key is that we assume these two investor subgroups have well-defined “habitats,” that they respond to differential sentiment or liquidity factors, and that arbitrage forces do not fully eliminate the effect of their trades.

Note also that the model does not assume individual investors are uniquely identified with noise trading. In fact, the model explicitly entertains the possibility that both institutional and retail investors are influenced by mass psychology, or differential time-varying liquidity needs. Thus, at least at the conceptual level, the sentiment or liquidity demands of both types of investors can influence returns. The focus of our paper, however, is on the sentiment of retail investors and its influence on stock returns.

---

<sup>10</sup> In fact, we provide evidence later that retail and institutional investors (i.e., those institutions required to file 13(f) forms) are not mirror images of each other. While retail investors tend to under-weight high institutionally-owned stocks, they do not necessarily over-weight the lowest institutionally-owned stocks.

### **III Data and Methodology**

#### *III.A Data*

The primary data for this study consist of trades and monthly portfolio positions of retail investors at a major discount brokerage house in the U.S. for the period of 1991-96. There are a total of 77,995 households in the database of which 62,387 have traded in stocks. The aggregate value of investor portfolios in our sample is, on average, 2.18 billion in a given month. An average investor holds a 4-stock portfolio (median is 3) with an average size of \$35,629 (median is \$13,869). Fewer than 10% of the investors hold portfolios over \$100,000 and fewer than 5% of them hold more than 10 stocks. The average monthly portfolio turnover rate which measures the frequency of trading is 6.59% (median is 2.53%) and a typical investor executes 9 trades per year where the average trade size is \$8,779 (median is \$5,239).<sup>11</sup>

Table I presents the summary statistics for the trading activities of investors in our sample. The individual investors in our sample execute 26,000 trades in a typical month and 1,244 trades on a typical day. As shown in Panel A, the number of buy trades in any given year is higher than the number of sell trades. In any given year, approximately 6000-7000 stocks are traded by the retail investors in our sample, suggesting that their trading spans a large set of stocks. However, as we show later, this activity is concentrated in firms with lower market capitalization. The stock and investor-level trading statistics (Panels B and C) show that in a typical month about 20% of stocks have 5 or more trades, and approximately 2-7% of these investors execute 5 or more trades.

In addition to the individual investor data, we also obtain quarterly institutional ownership data for the stocks in our sample from the CDA Spectrum database. Spectrum contains the end of

---

<sup>11</sup> See Barber and Odean (2000) for further details on the individual investor database.

quarter stock holdings of all institutions that file form 13F with the Securities and Exchange Commission (SEC). Institutions with more than \$100 million under management are required to file form 13F with the SEC and common stock positions of more than 10,000 shares or more than \$200,000 in value must be reported on the form. Using the quarterly institutional holdings, we compute the aggregate institutional ownership for each stock at the end of each quarter and this aggregate ownership data is used to construct ownership portfolios at the end of the year.

### *III.B Buy-Sell Imbalance Measure*

The aggregate trading activities of investors for a certain group of stocks (i.e., a stock portfolio) can be measured in a variety of ways. One such measure is a portfolio's buy-sell imbalance (BSI) during a certain time-period  $t$ . To compute the portfolio BSI, we first define the BSI for stock  $i$  in month  $t$  as

$$BSI_{it} = \frac{\sum_{j=1}^{D_t} (VB_{ijt} - VS_{ijt})}{\sum_{j=1}^{D_t} (VB_{ijt} + VS_{ijt})} \quad (6)$$

where

$D_t$  = number of days in month  $t$ ,

$VB_{ijt}$  = buy volume (measured in dollars) for stock  $i$  on day  $j$  in month  $t$ , and

$VS_{ijt}$  = sell volume (measured in dollars) for stock  $i$  on day  $j$  in month  $t$ .<sup>12</sup>

An equal-weighted average of individual stock BSIs provides a measure of portfolio-level BSI:

$$BSI_{pt} = \frac{100}{N_p} \sum_{i=1}^{N_p} BSI_{it} \quad (7)$$

---

<sup>12</sup> Our results are virtually unaffected when we use the number of shares instead of dollar volume to compute the monthly buy-sell imbalances.

where  $N_p$  is the number of stocks in portfolio  $p$ . Our measure of portfolio BSI gives the same weight to each stock in computing the portfolio sentiment measure. An alternative approach is to first compute the aggregate dollar volume in-flow (AVB) and aggregate dollar volume out-flow (AVS) for all the stocks in a portfolio, and define BSI as  $(AVB - AVS)/(AVB + AVS)$ . However, under this alternative approach, a single stock can strongly influence the portfolio BSI in a particular month, especially around information events such as earnings announcements and stock recommendation changes when the stock trading volume is unusually high. Our equal-weighted measure avoids such a bias.<sup>13</sup>

To remove the common dependence of portfolio BSI on the market factor, we perform the following regression:

$$BSI_{pt} = b_0 + b_1 RMRF_t + \varepsilon_{pt} \quad (8)$$

where  $BSI_{pt}$  is the buy-sell imbalance index for portfolio  $p$  in month  $t$ ,  $RMRF_t$  is the market return in excess of the riskfree rate in month  $t$ , and  $\varepsilon_{pt}$  is the residual BSI (or RBSI) for portfolio  $p$  in month  $t$ . In all our empirical analysis, we use this orthogonalized measure of investors' trading activities.

Table II reports the annual, aggregate BSI statistics for the raw BSI (Panel A) and residual BSI (Panel B). We find that the mean monthly raw BSI is positive during the first two years of our sample period and it is negative for the remaining four years. Furthermore, the marginal impact of the market factor on the aggregate BSI is not strong - the mean monthly residual BSI exhibits a pattern similar to the raw BSI.

---

<sup>13</sup> Most of the empirical analyses in this paper have been carried out using monthly aggregated measures of investor trading activities. Given the relatively sparse trading activities of individual investors in our sample, a daily buy-sell imbalance (BSI) measure for a particular stock is quite unreliable.

## IV Correlated Trading

The noise trader model asserts that the aggregate behavior of noise traders is affected by waves of common sentiment. In the absence of this type of systematic behavior, it is unlikely that noise trader sentiment can affect stock returns.<sup>14</sup> Our first set of tests is designed to examine the presence of a systematic component in the trading activities of retail investors. We examine the correlations in the trading activities of non-overlapping groups of investors as well as the correlations between pairs of non-overlapping stock portfolios.

### IV.A Correlations among Random Non-Overlapping Portfolios

To examine the correlations among stock portfolios, first, we form 1000 pairs of  $k$ -stock non-overlapping stock portfolios ( $k = 50, 75, 100, 125, 250, \text{ and } 500$ ) where stocks are chosen randomly from the set of stocks traded by investors in our sample. Next, for each of these randomly chosen portfolios, using equation (7), we construct a monthly buy-sell imbalance (BSI) time-series for the 71-month sample-period (January 1991 to November 1996). Then, using equation (8), we orthogonalize the raw BSI measures to obtain the residual BSI. Finally, we compute the correlations between the pairs of residual BSI indices of these non-overlapping portfolios and generate an empirical distribution of residual BSI correlation.

Table III reports the correlation statistics for different portfolio sizes. In Panel B, we present the residual BSI (or RBSI) correlation statistics and for comparison, we present the raw BSI correlation statistics in Panel A. In addition, Figure 1 shows an empirical distribution of these pair-wise correlations for  $k = 250$ . The average RBSI correlation is positive and significantly different from zero ( $p\text{-value} < 0.05$ ) for all chosen portfolio sizes. We find that the average RBSI correlation increases with portfolio size. For instance, for 250-stock portfolios, the average RBSI correlation is 0.600 while for 50-stock portfolios, this measure is only 0.234. A

---

<sup>14</sup> We interpret investor trading behavior as systematic if it is correlated across time and has strong cross-sectional correlation so that their trading activities do not “wash out” in the aggregate.

similar monotonic pattern is observed when we examine the correlation statistics of raw BSI. These results provide evidence of a systematic component in the trading activities of investors in our sample, which is uncorrelated with the movements of the market index.

To obtain an estimate of the average residual BSI correlation in the absence of a systematic component in the trading activities of the investors (i.e., a benchmark average residual BSI correlation) in our sample, we also conducted a Monte Carlo simulation. Specifically, we generate a BSI matrix where, for each stock, we keep the frequency and timing of trades fixed but randomly assigned a BSI in each month, where BSI lies uniformly in the range  $(-1,1)$ .<sup>15</sup> Using this simulated BSI matrix and following the procedure described earlier, we generate an empirical distribution of RBSI correlations by computing the correlations between 1000 pairs of non-overlapping 250-stock portfolios.<sup>16</sup> This entire process is repeated 500 times and a distribution of the average residual BSI correlation is obtained. We find that the average residual BSI correlation for the portfolios in this test is 0.04, which is much lower than the observed average residual BSI correlation of 0.60. In fact, we find that the average RBSI correlation is lower than 0.60 in each of our 500 repetitions. In short, the Monte Carlo results indicate that the observed RBSI correlations are unlikely to have been a chance occurrence ( $p$ -value  $< 0.01$ ).

#### *IV.B Correlations among Random Subsets of Investors*

For further evidence of systematic behavior, we examine the RBSI correlations between pairs of non-overlapping investor groups. Similar to our portfolio-based approach, we form 1000 pairs of non-overlapping investor groups each containing  $k$ -investors ( $k = 500, 1000, 1500, 2000, 2500, \text{ and } 5000$ ) where investors are chosen randomly from the set of investors in our sample. Next, for each of these randomly chosen investor groups, we construct a monthly buy-sell

---

<sup>15</sup> For robustness, we also conducted a similar test where, instead of choosing BSI randomly from  $(-1,1)$ , we re-sampled BSI from the observed empirical distribution of monthly BSI for each firm. In other words, we selected the monthly BSI randomly with replacement from a firm's population of monthly BSI measures. The results are qualitatively identical to the findings using a random draw from a uniform distribution.

<sup>16</sup> Other choices for the portfolio size (50, 100, and 500 stocks) produced similar results.

imbalance (BSI) time-series for the 71-month sample-period (January 1991 to November 1996).

The monthly BSI for investor group  $i$  in month  $t$  is obtained as:

$$BSI_{it} = \frac{\sum_{j=1}^{N_i} (VB_{jt} - VS_{jt})}{\sum_{j=1}^{N_i} (VB_{jt} + VS_{jt})} \quad (9)$$

where

$N_i$  = number of investors in group  $i$ ,

$VB_{jt}$  = buy volume (measured in dollars across all stocks) for investor  $j$  in month  $t$ , and

$VS_{jt}$  = sell volume (measured in dollars across all stocks) for investor  $j$  in month  $t$ .

Again, using equation (8), we orthogonalize the raw BSI measures to obtain the residual BSI. Finally, we compute the correlations between the pairs of residual BSI indices of these non-overlapping investor groups and generate an empirical distribution of residual BSI correlation.

Table IV reports the correlation statistics for different sizes of investor groups. In Panel B, we present the residual BSI (or RBSI) correlation statistics and for comparison, we present the raw BSI correlation statistics in Panel A. Analogous to our portfolio-based results, the average RBSI correlation is positive and statistically significant ( $p$ -value  $< 0.05$ ) for all chosen investor group sizes. Furthermore, the average RBSI correlation increases monotonically as the group size increases. For instance, for investor groups comprising of 1,000 investors each, the average RBSI correlation is 0.210 while for groups with 5,000 investors, the average RBSI correlation is 0.496. These results provide further evidence of systematic trading by retail investors in our sample.<sup>17</sup> Rather than becoming weaker as the number of investors increases, we find that the

---

<sup>17</sup> These findings are consistent with several recent, contemporaneous, studies. For example, Barber, Odean and Zhu (2003) provide evidence of correlated trading among individual investors, and explore psychology-based explanations for these patterns. Using a Chinese dataset, Feng and Seasholes (2002) find that the trading activities of investors that live within a certain geographic region are strongly correlated. In the same spirit, Jackson (2002)

correlation between average RBSI actually becomes stronger.

Overall, our results show that correlations between RBSI indices are strongly positive over non-overlapping stock portfolios and across non-overlapping groups of retail investors. Our randomization tests suggest these strong positive correlations could not have occurred by chance. Collectively, the evidence points to the existence of a systematic (or market-wide) component in the trading activities of retail investors. In particular, when retail investors buy (sell) one basket of stocks, they are likely to simultaneously buy (sell) other stock baskets.

## **V Characteristics of Retail Investor Sentiment**

### *V.A Construction of a Retail Investor Sentiment Factor*

In the second stage of our empirical tests, we examine the incremental explanatory power (over the four well-known risk factors) of a market-wide indicator of monthly shifts in retail investor sentiment across different baskets of stocks. We construct a retail investor sentiment factor (ISEN) which represents the difference between the returns on an equal-weighted portfolio of stocks with the most favorable retail sentiment (BSI deciles 8, 9, and 10) and the returns on an equal-weighted portfolio of stocks with the least favorable retail sentiment (BSI deciles 1, 2, and 3). To ensure that ISEN in a certain month can be constructed *ex ante*, we use the BSI in the *previous* month to measure the shift in retail sentiment and examine its influence on returns in the following month.

More specifically, the ISEN factor is constructed as follows: at the end of each month  $t$ , we sort all stocks in our sample using their BSI measures for month  $t$  and form ten decile portfolios. Portfolio 1 consists of stocks with the lowest BSI measure while portfolio 10 consists of stocks with the highest BSI measure. For each of the BSI decile portfolios constructed at the end of month  $t$ , we compute its return in month  $(t+1)$  as an equal weighted average of returns of stocks

---

provide additional evidence of systematic trading patterns among Australian investors. None of these studies link the correlated trading behavior to stock returns.

in that portfolio and a monthly return time-series is obtained for each of the ten BSI decile portfolios. Finally, a monthly ISEN time-series is computed, representing the spread between the average returns of portfolios 8, 9, and 10 and portfolios 1, 2, and 3.

#### *V.B Correlations among Sentiment Factor, Standard Risk Factors, and Macro-Variables*

As a starting point, we investigate the relation between our measure of retail investor sentiment and several empirically-inspired risk factors common in the literature. Specifically, Fama and French (1992) find that firms' market capitalization (size) and book-to-market ratio (B/M) explain a significant portion of the cross-sectional variation in stock returns. In a related study, Fama and French (1993) form mimicking portfolios based on size (SMB) and B/M (HML), and show that when the standard market model is augmented by these two variables, a number of pricing anomalies disappear. In this study, we examine the relation between ISEN, SMB, and HML and we also include a price momentum factor (UMD), as suggested by Jegadeesh and Titman (1993) and Carhart (1997).

Table V (Panel A) reports the basic statistics of the ISEN factor and four other standard risk factors (market or RMRF, small-minus-big or SMB, high-minus-low or HML, and momentum or UMD) and shows the correlation matrix. We find that the ISEN factor is weakly correlated with the standard risk factors - the contemporaneous correlations with RMRF, SMB, HML, and UMD factors are 0.115, 0.288, 0.117, and 0.063, respectively. Of these four correlations, only the correlation with SMB is statistically significant ( $p$ -value < 0.01).

We also examine whether retail investor sentiment is influenced by innovations in macro-economic variables that are likely to capture changes in market fundamentals (Chen, Roll and Ross (1986)). Specifically, we consider the following four macro-economic variables as potential determinants of investor sentiment: (i) *UI*: unexpected inflation where the average of 12 most recent inflation realizations is used to estimate the expected level of inflation, (ii) *MP*: monthly growth in industrial production, (iii)  $\Delta TS$ : change in the term spread where term spread

is the difference between the yield of a constant-maturity 10-year Treasury bond and the yield of a 3-month Treasury bill, and (iv)  $\Delta RP$ : change in the risk premium where the risk premium represents the difference between the yields of Moody's BAA-rated corporate bond and AAA-rated corporate bond. Table V (Panel B) shows that ISEN factor is only weakly correlated with these innovations in macro-economic variables - the contemporaneous correlations with UI, MP,  $\Delta TS$ , and  $\Delta RP$  are 0.191,  $-0.097$ ,  $-0.175$ , and  $-0.249$ , respectively. Of these four correlations, the correlation with  $\Delta RP$  is statistically significant at the 1% level while the correlation with UI is significant at the 10% level. The other two correlations are statistically insignificant.

Taken together, our correlation results suggest that changes in retail investor sentiment is unlikely to be fully explained by either the standard empirical risk factors (SMB, HML, and UMD) or innovations in key macro-economic variables (UI, MP,  $\Delta TS$ , and  $\Delta RP$ ). In the remaining part of the paper, we explore the possibility that changes in investor sentiment give rise to an orthogonal source of comovement in stock returns.

## **VI Retail Investor Sentiment and Stock Returns**

### *VI.A Characterizing the “Habitat” of Retail Investors*

Our model asserts that the impact of retail investor trading on returns will be most pronounced in stocks where these traders concentrate, i.e., their “habitat”. To identify the natural “habitat” of retail investors, we examine the concentration of stocks in their portfolio holdings, as well as the level of retail trading activity, across stock portfolios sorted by various firm characteristics. Table VI reports the results of these tests.

In Panel A, we report the mean percentage over (under) weight in retail ownership for firms sorted on Size, B/M, Institutional Ownership, and Price. To construct this panel, we compute a “benchmark” percentage retail ownership based on the total market capitalization of stocks that fall into each quintile portfolio at the end of each month. We then compute an “actual” percentage retail ownership based on the total market capitalization of stocks actually owned by

the retail investors in our database. Table values represent the difference between these two percentages, averaged across all months in our sample period.

Panel A results show that retail investors in our sample tend to concentrate their holdings in small, high B/M, low institutionally-owned, and low-priced stocks. For example, firms in the smallest size decile are over-weighted by 14.21% in the holdings of these retail investors, while firms in the largest size decile are under-weighted by 22.33%. The B/M results confirm earlier findings by Barber and Odean (2000) that retail investors, on average, exhibit a preference for value stocks. Interestingly, retail investors only slightly over-weight the lowest institutionally-owned stocks (1.19%), but are heavily under-weighted in the highest institutionally-owned stocks (−14.53%). Finally, retail investors also seem to avoid highly priced stocks (−30.27% under-weight).

We are interested not only in what stocks retail investors hold, but also how important their trades are in these stocks relative to the trades of other market participants. Panel B reports the results of a test of retail trading intensity. To measure the relative trading activity of retail investors in our sample, we obtain a normalized measure of retail trading activity (NTA) for each stock:

$$NTA_{it} = \frac{\text{Number of shares traded by investors in our sample}}{\text{Number of shares traded in the market}} \times 10^6. \quad (10)$$

$NTA_{it}$  is the normalized retail trading activity for stock  $i$  in month  $t$ . The NTA for a given portfolio is an equal-weighted average of stock-level NTA measures. Using the monthly NTA of each stock, we compute the average NTA for stock categories formed by sorting on size, book-to-market (B/M), stock price, and institutional ownership variables. The monthly averages of the NTA for the four stock categories provide additional indicators of retail investor “habitat.”

Panel B shows that the mean NTA decreases along size, institutional ownership, and stock price quintile portfolios, and it increases along the B/M quintile portfolio. In other words, retail trading volume is a larger proportion of the total market trading volume for small-cap stocks, value stocks, lower priced stocks, and stocks with lower institutional ownership. Consistent with the results from Panel B, the evidence from retail trading activity indicates that the natural “habitat” of the retail investor consists of small, low-priced, low institutionally-owned, and high B/M stocks. Consequently, we expect the influence of retail sentiment to be stronger among these stock categories.

#### *VI.B Estimation of Multi-Factor Time-Series Models: Size Sort*

To examine the incremental ability of ISEN to generate comovement in stock returns, our investigation follows procedures that have become standard in recent asset pricing studies. We employ a five-factor time-series model where the first three factors are from Fama and French (1993), the fourth factor is the momentum factor (e.g., Jegadeesh and Titman (1993), Carhart (1997)), and the fifth factor is our retail investor sentiment factor. The following factor model is used to test the sentiment-return time-series relation:

$$\begin{aligned}
 R_{pt} - R_{ft} = & \alpha_p + \beta_{1p} \text{RMRF}_t \\
 & + \beta_{2p} \text{SMB}_t + \beta_{3p} \text{HML}_t + \beta_{4p} \text{UMD}_t \\
 & + \beta_{5p} \text{ISEN}_t + \varepsilon_{pt}.
 \end{aligned} \tag{11}$$

Here,  $R_{pt}$  is the rate of return on the portfolio,  $R_{ft}$  is the riskfree rate of return,  $\text{RMRF}_t$  is the market return in excess of the riskfree rate,  $\text{SMB}_t$  is the difference between the value-weighted return of a portfolio of small stocks and the value-weighted return of a portfolio of large stocks,  $\text{HML}_t$  is the difference between the value-weighted return of a portfolio of high B/M stocks and the value-weighted return of a portfolio of low B/M stocks,  $\text{UMD}_t$  is the difference between the value-weighted return of a portfolio of stocks with high returns during months  $t-12$  to  $t-2$  and the value-weighted return of a portfolio of stocks with low returns during months  $t-12$  to  $t-2$ ,

$ISEN_t$  is the difference between the equal-weighted return of a portfolio of high BSI stocks and the equal-weighted return of a portfolio of low BSI stocks, and  $\varepsilon_{pt}$  is the residual return on the portfolio.

To set the stage, we consider portfolios obtained by sorting on firm size. Given our earlier findings on the habitat of retail investors, we expect the effect of retail investor sentiment to be most evident among small-cap stocks. At the end of each year, we sort the entire universe of stocks for which returns data is available from CRSP according to their market capitalizations at the end of November. Using NYSE break-points, we group stocks into size quintiles. Portfolio membership is not modified during the course of the year. Portfolio 1 consists of stocks with the lowest market capitalization while portfolio 5 contains stocks with the largest market capitalization. Next, for each portfolio, we compute the monthly portfolio return as an equal-weighted average of all stocks in the portfolio and construct a monthly portfolio return time-series. Finally, we estimate the time-series factor model (equation (11)) for each size quintile portfolio.

Table VII presents the time-series factor model estimates for each of the five size-quintile portfolios. For portfolio 1, the ISEN factor loading is positive (0.706) and statistically significant ( $t$ -value = 2.586). With a monthly standard deviation of 1.10% for the ISEN factor, one standard deviation shift in the sentiment measure corresponds to a 0.777% monthly shift in the portfolio return, which is economically significant. Similarly, for size portfolio 2, the ISEN factor loading is positive but considerably smaller in magnitude (0.162). Nevertheless, the coefficient estimate is statistically significant ( $t$ -value = 2.586). For the remaining 3 portfolios, the ISEN factor loadings are small in magnitude and statistically insignificant. These results suggest that stocks in lower size quintiles earn positive (negative) excess returns when retail investor sentiment grows more bullish (bearish) but retail sentiment does not influence the returns of other size quintile portfolios in a significant manner.

### *VI.C Other Habitat Measures: B/M, Institutional Ownership and Stock Price*

To further examine the variation in the degree of returns comovement generated by the retail investor sentiment factor (i.e., ISEN), we perform three additional one-dimensional sorts along B/M, institutional ownership (IO), and stock price dimensions. We carry out the five-factor model estimation for each set of portfolios. First, we examine the sentiment-return relation for five B/M-ranked portfolios. Analogous to the procedure for constructing size portfolios, at the end of each year, we sort the entire universe of stocks for which returns data is available from CRSP according to their B/M ratio at the end of December and assign each stock into a B/M quintile. The B/M break-points along with the portfolio break-points for other three sorting variables are reported in Table IX (Panel A).

We estimate the time-series factor model (equation (11)) for each B/M quintile portfolio. The estimation results are summarized in Table VIII (Panel B). We find that the ISEN factor loading is positive and statistically significant for high B/M stock portfolios (B/M portfolios 3, 4, and 5). This evidence supports BSW's "habitat" view of comovement - the systematic trading activities of retail investors are likely to be a part of the return generating process of value stocks where their concentration is higher. More specifically, the stocks in B/M quintile portfolios 3, 4, and 5 earn positive (negative) excess returns when retail investor sentiment grows more bullish (bearish).

We also estimate the time-series factor model (equation (11)) for each IO and price quintile portfolio. These results are also summarized in Table VIII (Panel B). As expected, the ISEN factor loading is strongly positive and significant for lower institutional ownership and lower price portfolios. Furthermore, it varies almost monotonically across the institutional ownership and stock price portfolios. For the lowest institutional ownership and price quintile portfolios, the sentiment factor loadings are 0.605 ( $p$ -value  $< 0.05$ ) and 0.934 ( $p$ -value  $< 0.01$ ), respectively.

The loadings decrease in magnitude and become statistically insignificant in higher IO and higher price quintile portfolios. This evidence suggests that low-institutional ownership and low priced firms earn positive (negative) excess returns when retail investor sentiment grows more bullish (bearish).

Our multi-factor model estimation results are consistent with the “habitat” view of comovement (Barberis, Shleifer, and Wurgler (2003)), which suggests the impact of shifts in retail investor sentiment on comovement in returns will be stronger among stocks where their concentration is higher. Specifically, when retail investors become more bullish (bearish), smaller stock, value stocks, low IO stocks, and lower priced stocks earn positive (negative) excess returns. In sum, the time-varying demands of retail investors induce comovement in returns even after controlling for the usual set of empirically-inspired risk factors.

#### *VI.D Arbitrage Costs and Sentiment-Return Relation*

Theory suggests that the sentiment-return relation is a function not only of investor habitat, but also of arbitrage costs. In a sentiment- or friction-based model of return comovements, investor sentiment has the greatest impact on the returns of firms that are the costliest to arbitrage. To examine the impact of arbitrage costs on the strength of sentiment-return relation, we estimate a time-series factor model (equation (11)) for four sets of portfolios formed by performing independent, double sorts on arbitrage cost and size, B/M, IO, and stock price variables. Following Wurgler and Zhurvaskaya (2002), we use the variance of the residual from a CAPM regression (i.e., the idiosyncratic risk of each firm) as a proxy for arbitrage cost. The monthly stock returns from the previous 60 months are used to estimate the CAPM regression.

The results are presented in Table IX. Panel A reports the mean arbitrage cost estimate for quintile portfolios sorted on firm size, B/M, Institutional Ownership, and Stock Price. As

expected, the average arbitrage costs are higher for stocks in the lower firm size, institutional ownership, and stock price quintiles. For these three sort variables, average arbitrage costs increase monotonically across the five quintiles. However, across the B/M portfolios, we find a U-shaped pattern, such that the arbitrage costs are high for both low B/M (i.e., growth) and high B/M (i.e., value) stock categories.

Of greater interest is the cross-sectional variation in the sentiment factor loadings when firms are sorted on arbitrage cost. Panel B reports the loading on the ISEN factor for stocks in five different arbitrage cost portfolios. Each row represents the results for a different firm characteristic. To hold the concentration of retail activity constant, we conduct this test using only stocks in the extreme quintile (i.e., firms that represent the natural habitat of retail investors). For example, row 1 reports the results for firms in the smallest size quintile. Similarly, rows 2 through 4 report results for the highest B/M, lowest IO, and lowest stock price quintiles. The average number of stocks in the portfolio is reported in parentheses next to the ISEN factor loading estimate.

Panel B shows that the ISEN factor loadings are stronger for higher arbitrage cost firms. This result is robust across all measures of retail habitat. For instance, within the lowest size quintile portfolio (row 1), the ISEN factor loadings increase across arbitrage cost quintile portfolios. In fact, when arbitrage costs are low, the ISEN factor loading is insignificant even for small firms. We observe a similar pattern across the other three rows. In each instance, when arbitrage costs are higher, the loading on ISEN is stronger. Consistent with a sentiment-based model of return comovements, these results suggest that arbitrage cost and retail investor habitat are joint determinants of the strength of sentiment-return relation.<sup>18</sup>

---

<sup>18</sup> In unreported results, we find that the ISEN loadings are insignificant for portfolios with higher arbitrage costs when the stocks in those portfolios do not belong to the retail habitat. For instance, growth stocks do not have a strong retail clientele and we find that growth stock portfolios have insignificant ISEN loadings even when arbitrage costs are high.

## **VII Summary and Conclusions**

In this study, using a large dataset of retail trades from a major discount brokerage house in the U.S., we examine the effect of retail trading patterns on stock returns. Our results show that the trading activities of retail investors contain a systematic component that is uncorrelated with the overall market movements. Using an aggregate measure of retail investors' trading activities, we construct a market-wide retail investor sentiment factor (ISEN), representing the spread between the returns on an equal-weighted portfolio of stocks with high buy-sell imbalance and the returns on an equal-weighted portfolio of stocks with low buy-sell imbalance.

We find that this measure of retail investor sentiment has incremental explanatory power (over the standard risk factors) for small stocks, value stocks, stocks with low institutional ownership, and stocks with lower prices. The direction of the relation indicates that when retail investors grow relatively bullish, the stocks in these portfolios enjoy higher excess returns. The magnitude of their influence is affected by factors associated with retail investor habitat and cross-sectional differences in arbitrage costs.

Specifically, we show that retail investors concentrate their holdings and their trading activities in smaller, lower-priced, higher B/M, and lower institutionally-owned firms. At the same time, we find that these are the firms most sensitive to changes in retail investor sentiment. Moreover, controlling for retail investor concentration, we show that firms with higher arbitrage costs (i.e., higher idiosyncratic risk) exhibit much stronger sensitivity to the ISEN factor.

Collectively, these findings are broadly consistent with the predictions of noise trader models in which the systematic activities of retail investors affect the returns of those stocks in which they are concentrated. In particular, our results provide support for a sentiment-based theory of returns comovement advanced by Barberis, Shleifer, and Wurgler (BSW; 2003). Consistent with the "habitat" version of the BSW model, we find that stocks preferred by retail investors are the

ones most sensitive to the market-wide retail investor sentiment factor (i.e., ISEN). Also consistent with their model, we find that the strength of the sentiment-return relation is a function of arbitrage costs.

More broadly, our results are suggestive of a role for mass psychology in the study of financial markets. The traditional case against a role for mass psychology in markets is based on two key assertions: (1) the cognitive foibles committed by individuals do not aggregate across the investing populous (i.e., individual irrationalities do not result in systematic directional behavior across large groups of investors); and (2) even if systematic noise trading exists, an army of rational arbitrageurs stands ready to offset this behavior, leaving prices unaffected. Our results suggest that, at least in the case of retail investors, both assertions are questionable.

Our findings also raise a number of interesting issues for future research. In particular, questions remain as to the factors that affect the proclivity of individual investors to buy or sell stocks - i.e., whether their time-varying preference is due to liquidity concerns, risk aversion, or irrational sentiment. Indeed, our findings highlight the need to better understand the processes by which individual investors formulate their trading decisions, including an identification of the information sources they use in decision making, and the nature of their belief updating process. We hope to address some of these topics in our on-going research.

## References

Baker, Malcolm, and Jeffrey Wurgler, 2003, Investor sentiment and the cross-section of stock returns, Working paper, Harvard Business School and NYU Stern School, November 2003.

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

Barber, Brad, and Terrance Odean, 2001, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, Working Paper, Haas School of Business, University of California at Berkeley, November 2001.

Barber, Brad, and Terrance Odean, and Ning Zhu, 2003, Systematic noise, Working Paper, Haas School of Business, University of California at Berkeley, April 2003.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2003, Comovement, *Journal of Financial Economics* Forthcoming.

Battalio, Robert H., and Richard R. Mendenhall, 2003, Earnings Expectations and Investor Clienteles, Working Paper, Mendoza College of Business, University of Notre Dame, December 2003.

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

Bodurtha, James N., Dong-Soon Kim, and Charles M.C. Lee, 1995, Closed-end Country Funds and U.S. Market Sentiment, *Review of Financial Studies* 8, 879-918.

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

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383-403.

Chen, Nai-Fu, Raymond Kan, and Merton H. Miller, 1993a, Are the discounts on closed-end funds a sentiment index?, *Journal of Finance* 48, 785-800.

Chen, Nai-Fu, Raymond Kan, and Merton H. Miller, 1993b, A Rejoinder, *Journal of Finance* 48, 809-810.

Chopra, Navin, Charles M.C. Lee, Andrei Shleifer, and Richard H. Thaler, 1993a, Yes, discounts

on closed-end funds are a sentiment index, *Journal of Finance* 48, 801-808.

Chopra, Navin, Charles M.C. Lee, Andrei Shleifer, and Richard H. Thaler, 1993b, Summing Up, *Journal of Finance* 48, 811-812.

DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703-738.

Elton, Edward J., Martin J. Gruber, and Jeffrey A. Busse, 1998, Do investors care about sentiment?, *Journal of Business* 77, 477-501.

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

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

Feng, Lei, and Mark S. Seasholes, 2003, Correlated trading and location, *Journal of Finance* Forthcoming.

Frieder, Laura and Avanidhar Subrahmanyam, 2002, Brand perceptions and the market for common stock, Working Paper, UCLA.

Gemmill, Gordon, and Dylan C. Thomas, 2002, Noise-trading, costly arbitrage, and asset prices: Evidence from closed end funds, *Journal of Finance* 57, 2571-2594.

Jackson, Andrew, 2003, The Aggregate Behavior of Individual Investors, Working Paper, London Business School, July 2003.

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

Lamont, Owen A., and Richard H. Thaler, 2002, Can the market add and subtract? Mispricing in tech stock carve-outs, Working Paper, Graduate School of Business, University of Chicago, June 2002.

Lease, Ronald C., Wilbur G. Lewellen, and Gary G. Schlarbaum, 1974, The individual investor: Attributes and attitudes, *Journal of Business* 29, 413-433.

Lee, Charles M.C., Andrei Shleifer, and Richard H. Thaler, 1991, Investor sentiment and the

closed-end fund puzzle, *Journal of Finance* 46, 75-108.

Lee, Charles M. C., 2001, Market efficiency and Accounting Research: A Discussion, *Journal of Accounting and Economics* 31, 2001, 233-253.

Lee, Charles M. C., 1992, Earnings News and Small Traders: An Intraday Analysis, *Journal of Accounting and Economics* 15, 265-302.

Lewellen, Wilbur G., Ronald C. Lease, and Gary G. Schlarbaum, 1977, Patterns of investment strategy and behavior among individual investors, *Journal of Business* 50, 296-333.

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

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

Rashes, Michael S., 2001, Massively confused investors making conspicuously ignorant choices, *Journal of Finance* 56, 1911-1927.

Shiller, Robert J., 1984, Stock prices and social dynamics, *Brookings Papers on Economic Activity* 2, 457-510.

Shiller, Robert J., and John Pound, 1989, Survey evidence on diffusion of interest and information among investors, *Journal of Economic Behavior and Organization* 12, 47-66.

Shleifer, Andrei, 2000, *Inefficient Markets*, Oxford University Press Inc., New York, NY.

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35-55.

Wurgler, Jeffrey, and Ekaterina Zhuravskaya, 2002, Does Arbitrage Flatten Demand Curve for Stocks?, *Journal of Business* 75, 583-608.

Yunker, James A., and Timothy L. Krehbiel, 1974, Investment analysis by the individual investor, *Quarterly Review of Economics and Business* 28, 90-101.

**Table I**  
**Summary Statistics: Investor Trading Behavior**

This table reports aggregate, stock-level, and investor-level trading statistics. There are 62,387 individual investors in our sample who executed 1,854,776 trades in 10,877 stocks during the 1991-96 sample-period. The statistics are reported only for trades for which returns data are available from CRSP and are used in our study. In Panel A, we report various aggregate trading statistics, in Panel B, we report stock-level trading statistics, and in Panel C, we report investor-level statistics. The number of stocks and the number of investors with a valid stock position at the end of the most recent month prior to a trade are used to obtain the proportions in Panels B and C respectively. The individual investor holdings data are from a large discount brokerage house in the U.S. for the 1991-96 time-period.

*Panel A: Aggregate Trading Statistics*

Statistic	1991	1992	1993	1994	1995	1996
<i>Number of buy trades</i>	184,358	174,466	168,790	141,570	167,134	179,417
<i>Number of sell trades</i>	133,845	132,541	146,768	120,209	151,427	154,251
<i>Average trade size (buys)</i>	\$12,941	\$13,166	\$13,114	\$12,698	\$13,004	\$12,815
<i>Average trade size (sells)</i>	\$13,147	\$12,927	\$12,953	\$12,977	\$13,043	\$12,853
<i>Total number of stocks traded</i>	5,930	6,010	6,514	6,804	7,206	7,096
<i>Total number of investors</i>	42,109	41,222	39,712	35,850	35,800	34,262

*Panel B: Monthly Stock-Level Trading Statistics: Proportion of Stocks (in Percent)*

Number of Monthly Trades	1991	1992	1993	1994	1995	1996
<i>At least 1 trade</i>	56.94	54.16	57.85	55.98	58.63	60.29
<i>5 or more trades</i>	19.75	17.52	18.34	15.81	17.66	18.12
<i>10 or more trades</i>	10.36	9.21	8.77	7.26	8.53	8.98
<i>25 or more trades</i>	3.60	3.25	2.92	2.37	2.87	3.25
<i>50 or more trades</i>	1.45	1.32	1.22	0.89	1.07	1.38
<i>75 or more trades</i>	0.86	0.77	0.71	0.49	0.59	0.82
<i>100 or more trades</i>	0.55	0.49	0.47	0.32	0.40	0.56

*Panel C: Monthly Investor-Level Trading Statistics: Proportion of Investors (in Percent)*

Number of Monthly Trades	1991	1992	1993	1994	1995	1996
<i>At least 1 trade</i>	21.91	20.00	23.17	27.06	36.61	44.72
<i>5 or more trades</i>	2.33	2.04	2.58	2.92	5.03	6.99
<i>10 or more trades</i>	0.56	0.47	0.60	0.68	1.37	2.17
<i>25 or more trades</i>	0.22	0.18	0.23	0.26	0.56	1.01
<i>50 or more trades</i>	0.11	0.09	0.11	0.12	0.29	0.55
<i>75 or more trades</i>	0.06	0.05	0.06	0.07	0.17	0.34
<i>100 or more trades</i>	0.01	0.01	0.01	0.02	0.03	0.07

**Table II**  
**Aggregate Buy-Sell Imbalance Statistics**

This table reports the summary statistics of the aggregate buy-sell imbalance (BSI) time-series. The aggregate BSI in month  $t$  is defined as,  $BSI_t = \frac{100}{N_t} \sum_{i=1}^{N_t} BSI_{it}$  where the BSI for stock  $i$  in month  $t$  is defined as  $BSI_{it} = [\sum_{j=1}^{D_t} (VB_{ijt} - VS_{ijt})] / [\sum_{j=1}^{D_t} (VB_{ijt} + VS_{ijt})]$ . Here,  $D_t$  is the number of days in month  $t$ ,  $VB_{ijt}$  is the buy volume (measured in dollars) for stock  $i$  on day  $j$  in month  $t$ ,  $VS_{ijt}$  is the sell volume (measured in dollars) for stock  $i$  on day  $j$  in month  $t$ , and  $N_t$  is the total number of stocks traded by investors in our sample in month  $t$ . The residual BSI is obtained by removing the common dependence of aggregate BSI on the market. The individual investor holdings data are from a large discount brokerage house in the U.S. for the 1991-96 time-period.

*Panel A: Raw Buy-Sell Imbalance (in Percent)*

Statistic	1991	1992	1993	1994	1995	1996
<i>Mean</i>	5.47	2.96	-3.79	-0.08	-5.49	-3.57
<i>Median</i>	6.65	3.99	-2.20	0.04	-5.97	-2.47
<i>Std. Dev.</i>	7.13	6.62	3.57	6.51	6.08	3.57
<i>Minimum</i>	-15.55	-12.41	-12.12	-16.18	-20.06	-10.02
<i>Maximum</i>	12.35	12.73	0.23	7.63	4.51	0.70

*Panel B: Residual Buy-Sell Imbalance (in Percent)*

Statistic	1991	1992	1993	1994	1995	1996
<i>Mean</i>	6.98	3.15	-3.46	-0.44	-3.93	-2.51
<i>Median</i>	9.00	5.37	-3.83	-0.06	-3.39	-2.35
<i>Std. Dev.</i>	5.76	6.31	3.17	6.21	6.05	3.49
<i>Minimum</i>	-8.01	-11.47	-11.04	-15.59	-19.28	-8.34
<i>Maximum</i>	12.38	10.96	0.79	7.98	5.67	2.42

**Table III**  
**Evidence of Systematic Trading using Random Stock Portfolios**

This table reports the simulation results (correlation statistics) from our first set of randomization tests that examines the existence of a systematic component in investors' trading activities using random portfolios. 1,000 pairs of non-overlapping  $k$ -stock portfolios are formed where  $k = 50, 75, 100, 125, 250$ , and 500. The buy-sell imbalance (BSI) time-series is obtained for these portfolios where BSI for portfolio  $p$  in month  $t$  is defined as,  $BSI_{pt} = \frac{100}{N_{pt}} \sum_{i=1}^{N_{pt}} BSI_{it}$  where the BSI for stock  $i$  in month  $t$  is defined as  $BSI_{it} = [\sum_{j=1}^{D_t} (VB_{ijt} - VS_{ijt})] / [(\sum_{j=1}^{D_t} (VB_{ijt} + VS_{ijt}))]$ . Here,  $D_t$  is the number of days in month  $t$ ,  $VB_{ijt}$  is the buy volume (measured in dollars) for stock  $i$  on day  $j$  in month  $t$ ,  $VS_{ijt}$  is the sell volume (measured in dollars) for stock  $i$  on day  $j$  in month  $t$ , and  $N_{pt}$  is the number of stocks in portfolio  $p$  formed in month  $t$ . The residual BSI (RBSI) is obtained by removing the common dependence of BSI on the market and finally, the correlations between pairs of RBSI time-series are computed. In Panel A, we report the raw BSI correlation statistics while in Panel B we report the residual BSI correlation statistics. The individual investor holdings data are from a large discount brokerage house in the U.S. for the 1991-96 time-period.

*Panel A: Raw Buy-Sell Imbalance Correlation Statistics*

Statistic	Number of stocks in the random portfolio					
	50	75	100	125	250	500
<i>Mean</i>	0.263	0.407	0.505	0.576	0.638	0.628
<i>Median</i>	0.265	0.408	0.509	0.583	0.642	0.631
<i>Std. Dev.</i>	0.123	0.097	0.086	0.073	0.062	0.062
<i>25<sup>th</sup> Pctl.</i>	0.176	0.342	0.448	0.533	0.599	0.586
<i>75<sup>th</sup> Pctl.</i>	0.344	0.476	0.564	0.625	0.680	0.674
<i>Minimum</i>	-0.083	0.106	0.182	0.369	0.380	0.426
<i>Maximum</i>	0.578	0.677	0.737	0.789	0.782	0.794

*Panel B: Residual Buy-Sell Imbalance Correlation Statistics*

Statistic	Number of stocks in the random portfolio					
	50	75	100	125	250	500
<i>Mean</i>	0.234	0.367	0.461	0.534	0.600	0.612
<i>Median</i>	0.237	0.372	0.464	0.543	0.605	0.618
<i>Std. Dev.</i>	0.127	0.104	0.093	0.080	0.069	0.065
<i>25<sup>th</sup> Pctl.</i>	0.142	0.296	0.399	0.484	0.556	0.566
<i>75<sup>th</sup> Pctl.</i>	0.317	0.441	0.524	0.588	0.648	0.659
<i>Minimum</i>	-0.127	0.026	0.142	0.311	0.316	0.394
<i>Maximum</i>	0.548	0.656	0.719	0.720	0.768	0.789

**Table IV**  
**Evidence of Systematic Trading using Random Subsets of Investors**

This table reports the simulation results (correlation statistics) from our second set of randomization tests that examines the existence of a systematic component in investors' trading activities using randomly chosen sets of investors. 1,000 pairs of mutually exclusive  $k$ -investor sets are formed where  $k = 500, 1000, 1500, 2000, 2500,$  and  $5000$ . The buy-sell imbalance (BSI) time-series is obtained for these investor groups where BSI for group  $i$  in month  $t$  is defined as  $BSI_{it} = [\sum_{j=1}^{N_i} (VB_{jt} - VS_{jt})] / [(\sum_{j=1}^{N_i} (VB_{jt} + VS_{jt}))]$ . Here,  $N_i$  is the number of investors in group  $i$ ,  $VB_{jt}$  is the buy volume (measured in dollars across all stocks) for investor  $j$  in month  $t$ , and  $VS_{jt}$  is the sell volume (measured in dollars across all stocks) for investor  $j$  in month  $t$ . The residual BSI (RBSI) is obtained by removing the common dependence of BSI on the market and finally, the correlations between pairs of RBSI time-series are computed. In Panel A, we report the raw BSI correlation statistics while in Panel B we report the residual BSI correlation statistics. The individual investor holdings data are from a large discount brokerage house in the U.S. for the 1991-96 time-period.

*Panel A: Raw Buy-Sell Imbalance Correlation Statistics*

Statistic	Number of randomly chosen investors					
	500	1000	1500	2000	2500	5000
<i>Mean</i>	0.188	0.262	0.333	0.390	0.426	0.562
<i>Median</i>	0.180	0.263	0.334	0.403	0.426	0.575
<i>Std. Dev.</i>	0.133	0.114	0.126	0.118	0.114	0.109
<i>25<sup>th</sup> Pctl.</i>	0.097	0.182	0.251	0.323	0.375	0.492
<i>75<sup>th</sup> Pctl.</i>	0.293	0.348	0.413	0.463	0.504	0.646
<i>Minimum</i>	-0.188	-0.046	0.046	0.042	0.044	0.262
<i>Maximum</i>	0.542	0.536	0.651	0.671	0.704	0.786

*Panel B: Residual Buy-Sell Imbalance Correlation Statistics*

Statistic	Number of randomly chosen investors					
	500	1000	1500	2000	2500	5000
<i>Mean</i>	0.147	0.210	0.272	0.326	0.365	0.496
<i>Median</i>	0.141	0.209	0.278	0.339	0.367	0.506
<i>Std. Dev.</i>	0.127	0.112	0.124	0.117	0.110	0.108
<i>25<sup>th</sup> Pctl.</i>	0.068	0.133	0.188	0.252	0.295	0.436
<i>75<sup>th</sup> Pctl.</i>	0.240	0.295	0.352	0.398	0.434	0.582
<i>Minimum</i>	-0.172	-0.105	-0.028	-0.067	-0.043	0.167
<i>Maximum</i>	0.538	0.484	0.591	0.602	0.651	0.763

**Table V**  
**Correlations Among the Standard Risk Factors, Macro-Economic Variables,**  
**and the Individual Investor Sentiment Factor**

This table reports the correlations among the standard risk factors and the individual investor sentiment factor (Panel A), and macro-economic variables and the individual investor sentiment factor (Panel B) during the 1991-96 sample period. RMRF is the market return in excess of the riskfree rate, SMB is the difference between the value-weighted return of a portfolio of small stocks and the value-weighted return of a portfolio of large stocks, HML is the difference between the value-weighted return of a portfolio of high B/M stocks and the value-weighted return of a portfolio of low B/M stocks, and UMD is the difference between the value-weighted return of a portfolio of stocks with high returns during months  $t - 12$  to  $t - 2$  and the value-weighted return of a portfolio of stocks with low returns during months  $t - 12$  to  $t - 2$ . Four macro-economic variables are used: UI is the unexpected inflation where the average of 12 most recent inflation realizations is used to estimate the expected level of inflation, MP is the monthly growth in industrial production,  $\Delta TS$  is the change in the term-spread (TS) where TS, a measure of the term structure, is the difference between the yield of a constant-maturity 10-year Treasury bond and the yield of a 3-month Treasury bill, and  $\Delta RP$  is the change in the risk-premium where RP, a measure of the risk premium, represents the difference between the yields of Moody's BAA-rated corporate bond and AAA-rated corporate bond. The individual investor sentiment factor ISEN is the difference between the equal-weighted return of a portfolio of stocks with high buy-sell imbalance (deciles 8, 9, and 10) and the equal-weighted return of a portfolio of stocks with low buy-sell imbalance (deciles 1, 2, and 3). The individual investor holdings data are from a large discount brokerage house in the U.S. for the 1991-96 time-period.

*Panel A: Correlations Among the Standard Risk Factors and the Investor Sentiment Factor*

Factor	Mean (%)	StdDev (%)	RMRF	SMB	HML	UMD
<i>Market (RMRF)</i>	1.00	2.86				
<i>Small-Minus-Big (SMB)</i>	0.17	2.46	0.070			
<i>High-Minus-Low (HML)</i>	0.45	2.42	-0.374	-0.312		
<i>Momentum (UMD)</i>	0.88	2.32	0.206	-0.047	-0.075	
<i>Sentiment (ISEN)</i>	0.08	1.10	0.115	0.288	0.117	0.063

*Panel B: Correlations Among Innovations in Macro Variables and the Investor Sentiment Factor*

Factor	Mean (%)	StdDev (%)	UI	MP	$\Delta RP$	$\Delta TS$
<i>Unexpected Inflation (UI)</i>	-0.02	0.12				
<i>Growth in Monthly Prod. (MP)</i>	0.31	0.49	0.223			
<i>Change in Risk Premium (<math>\Delta RP</math>)</i>	-1.08	6.63	-0.118	0.150		
<i>Change in Term Spread (<math>\Delta TS</math>)</i>	0.79	2.98	0.071	0.281	-0.014	
<i>Sentiment (ISEN)</i>	0.08	1.10	0.191	-0.097	-0.175	-0.249

**Table VI**  
**Ownership Concentration and Trading Activity of Retail Investors**

This table reports the ownership concentration and trading activity of retail investors across stocks sorted by various firm characteristics. In Panel A, we report the mean percentage over (under) weight in retail ownership for firms sorted on Size, B/M, Institutional Ownership, and Price. To construct this panel, we compute a “benchmark” percentage ownership based on the total market capitalization of stocks that fall into each quintile portfolio at the end of each month. We then compute an “actual” percentage retail ownership based on the total market capitalization of stocks actually owned by the retail investors in our sample. Panel values represent the difference between these two percentages, averaged across all months in our sample period. In Panel B, we report the relative trading activity for different stock categories (or styles). The relative trading activity (or concentration) of retail investors in our sample is defined as:

$$NTA_{it} = \frac{\text{Number of shares traded by investors in our sample}}{\text{Number of shares traded in the market}} \times 10^6$$

where  $NTA_{it}$  is the normalized retail trading activity for stock  $i$  in month  $t$ . Using the monthly NTA, we compute the average NTA for stock categories formed by sorting on size, book-to-market (B/M), stock price, and institutional ownership variables. The individual investor holdings data are from a large discount brokerage house in the U.S. for the 1991-96 time-period.

*Panel A: Mean Percentage Over (Under) Weight  
in Retail Ownership*

Quintile	Stock Category			
	Size	B/M	InstiOwn	Price
<i>Q1 (Low)</i>	14.21	-5.26	1.19	9.59
<i>Q2</i>	5.14	-3.98	4.01	8.77
<i>Q3</i>	2.69	-1.78	5.64	6.19
<i>Q4</i>	0.29	2.34	3.68	5.67
<i>Q5 (High)</i>	-22.33	8.67	-14.53	-30.27

*Panel B: Mean Retail Trading Activity (NTA)*

Quintile	Stock Category			
	Size	B/M	InstiOwn	Price
<i>Q1 (Low)</i>	3.37	0.29	3.01	2.03
<i>Q2</i>	0.48	0.47	2.70	1.35
<i>Q3</i>	0.71	2.04	2.50	1.44
<i>Q4</i>	0.32	1.38	1.39	1.01
<i>Q5 (High)</i>	0.69	1.80	0.51	0.65

**Table VII**  
**Time-Series Factor Model Estimates for Size Portfolios**

This table reports the factor model estimates for the 5 size-quintile portfolios. The quintile portfolios are formed at the end of each year in December using NYSE size break-points and then held fixed throughout the following year. The following time-series factor model is estimated:

$$R_{pt} - R_{ft} = \alpha_p + \beta_{1p}RMRF_t + \beta_{2p}SMB_t + \beta_{3p}HML_t + \beta_{4p}UMD_t + \beta_{5p}ISEN_t + \varepsilon_{pt} \quad t = 1, 2, \dots, T.$$

Here,  $R_{pt}$  is the rate of return on the size portfolio,  $R_{ft}$  is the riskfree rate of return,  $RMRF_t$  is the market return in excess of the riskfree rate,  $SMB_t$  is the difference between the value-weighted return of a portfolio of small stocks and the value-weighted return of a portfolio of large stocks,  $HML_t$  is the difference between the value-weighted return of a portfolio of high B/M stocks and the value-weighted return of a portfolio of low B/M stocks,  $UMD_t$  is the difference between the value-weighted return of a portfolio of stocks with high returns during months  $t - 12$  to  $t - 2$  and the value-weighted return of a portfolio of stocks with low returns during months  $t - 12$  to  $t - 2$ ,  $ISEN_t$  is the difference between the equal-weighted return of a portfolio of stocks with high buy-sell imbalance (top three deciles) and the equal-weighted return of a portfolio of stocks with low buy-sell imbalance (bottom three deciles) in month  $t$ , and  $\varepsilon_{pt}$  is the residual return on the portfolio. The individual investor holdings data are from a large discount brokerage house in the U.S. for the 1991-96 time-period. The Newey-West adjusted  $t$ -values of the coefficient estimates are reported in the parentheses.

Portfolio	Alpha	RMRF	SMB	HML	UMD	ISEN	Adj. $R^2$
<i>Size</i>	0.371	0.896	1.398	0.681	-0.252		
<i>Quintile 1</i>	(1.475)	(10.243)	(8.687)	(5.681)	(-2.923)		0.790
	0.363	0.854	1.307	0.714	-0.214	<b>0.706</b>	
	(1.382)	(9.445)	(9.364)	(7.040)	(-2.375)	<b>(2.586)</b>	0.808
<i>Size</i>	-0.054	0.995	0.922	0.187	-0.166		
<i>Quintile 2</i>	(-0.495)	(12.479)	(23.317)	(4.945)	(-4.161)		0.948
	-0.089	0.973	0.892	0.196	-0.127	<b>0.162</b>	
	(-0.826)	(22.647)	(10.775)	(5.312)	(-3.433)	<b>(2.164)</b>	0.957
<i>Size</i>	-0.143	1.017	0.734	0.132	-0.017		
<i>Quintile 3</i>	(-1.332)	(17.512)	(15.323)	(5.938)	(-0.824)		0.974
	-0.174	1.005	0.725	0.133	0.011	0.011	
	(-1.187)	(18.393)	(17.102)	(6.565)	(0.605)	(0.166)	0.974
<i>Size</i>	-0.014	1.032	0.427	0.104	-0.038		
<i>Quintile 4</i>	(-0.225)	(12.135)	(14.747)	(5.630)	(-1.819)		0.959
	-0.025	1.028	0.426	0.104	-0.028	-0.015	
	(-0.386)	(9.116)	(12.175)	(5.751)	(-1.172)	(-0.208)	0.957
<i>Size</i>	0.006	1.055	0.007	0.066	-0.062		
<i>Quintile 5</i>	(0.163)	(17.201)	(0.421)	(3.110)	(-2.924)		0.982
	-0.006	1.053	0.011	0.064	-0.053	-0.059	
	(-0.155)	(14.116)	(0.578)	(3.226)	(-1.977)	(-1.295)	0.983

**Table VIII**  
**Sentiment Factor Loadings for Portfolios formed on Firm Characteristics**

This table reports the sentiment factor loadings for portfolios formed on firm size, book-to-market (B/M), level of institutional ownership (IO), and month-end stock price. The quintile portfolios are formed at the end of each year using the size, B/M, IO, and price break-points from the end of December. The portfolios are held fixed throughout the following year. Panel A reports the portfolio break-points while Panel B reports the ISEN factor loadings. The following time-series factor model is estimated:

$$R_{pt} - R_{ft} = \alpha_p + \beta_{1p}RMRF_t + \beta_{2p}SMB_t + \beta_{3p}HML_t + \beta_{4p}UMD_t + \beta_{5p}ISEN_t + \varepsilon_{pt} \quad t = 1, 2, \dots, T.$$

Here,  $R_{pt}$  is the rate of return on portfolio  $p$ ,  $R_{ft}$  is the riskfree rate of return,  $RMRF_t$  is the market return in excess of the riskfree rate,  $SMB_t$  is the difference between the value-weighted return of a portfolio of small stocks and the value-weighted return of a portfolio of large stocks,  $HML_t$  is the difference between the value-weighted return of a portfolio of high B/M stocks and the value-weighted return of a portfolio of low B/M stocks,  $UMD_t$  is the difference between the value-weighted return of a portfolio of stocks with high returns during months  $t - 12$  to  $t - 2$  and the value-weighted return of a portfolio of stocks with low returns during months  $t - 12$  to  $t - 2$ ,  $ISEN_t$  is the difference between the equal-weighted return of a portfolio of stocks with high buy-sell imbalance (top three deciles) and the equal-weighted return of a portfolio of stocks with low buy-sell imbalance (bottom three deciles) in month  $t$ , and  $\varepsilon_{pt}$  is the residual return on the portfolio. The individual investor holdings data are from a large discount brokerage house in the U.S. for the 1991-96 time-period. \* and \*\* denote significance at the 5% and 1% levels, respectively.

*Panel A: Portfolio Break-Points*

Sorting Variable	Quintile Portfolios				
	Low	Q2	Q3	Q4	High
<i>Firm Size (in millions)</i>	< 144	144-401	401-981	981-2,738	> 2,738
<i>Book-To-Market</i>	< 0.365	0.365-0.569	0.569-0.783	0.783-1.077	> 1.077
<i>Insti. Own. (in percent)</i>	< 4.10	4.10-14.14	14.14-28.67	28.67-49.40	> 49.40
<i>Stock Price</i>	<3.55	3.55-8.25	8.25-14.22	14.22-24.24	>24.24

*Panel B: Sentiment Factor Loadings*

Sorting Variable	Quintile Portfolios				
	Low	Q2	Q3	Q4	High
<i>Firm Size</i>	0.706**	0.162**	0.011	-0.015	-0.059
<i>Book-To-Market</i>	0.308	0.450	0.282*	0.202**	0.595**
<i>Insti. Own.</i>	0.605*	0.751**	0.389	0.183	0.036
<i>Stock Price</i>	0.934**	0.736**	0.364	0.085	0.001

**Table IX**  
**Arbitrage Costs and Retail Sentiment**

This table examines the effect of arbitrage costs on the strength of sentiment-return relation. We use the idiosyncratic risk of a stock, defined as the variance of the residual from the market return regression model, as a measure of arbitrage cost. The market model is estimated using monthly returns data over the past 60 months. To focus on the effect of arbitrage costs, we use only firms with significant retail concentration - i.e., firms in the lowest quintile by Size, IO, and Price, and highest quintile by B/M. Quintile portfolios are formed at the end of each year using break-points from the end of December. Panel A reports the mean idiosyncratic risk for quintile portfolios sorted by Size, B/M, Institutional Ownership (IO), and Price. Panel B reports the sentiment factor loadings for quintile portfolios sorted by arbitrage cost. The average number of stocks in a portfolio is reported in the parenthesis. To obtain the sentiment factor loading for a portfolio, we estimate the following time-series factor model:

$$R_{pt} - R_{ft} = \alpha_p + \beta_{1p}RMRF_t + \beta_{2p}SMB_t + \beta_{3p}HML_t + \beta_{4p}UMD_t + \beta_{5p}ISEN_t + \varepsilon_{pt} \quad t = 1, 2, \dots, T.$$

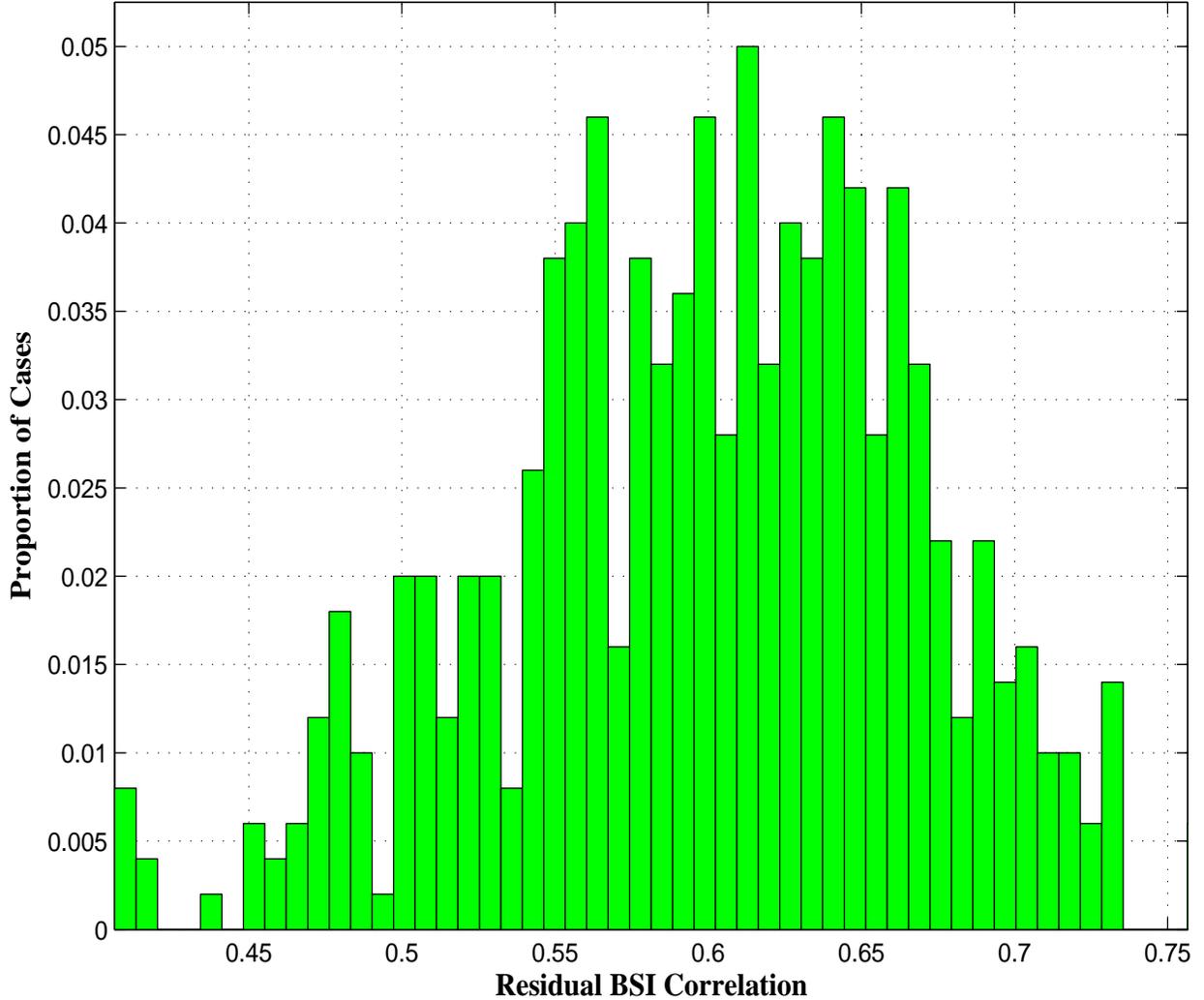
Here,  $R_{pt}$  is the rate of return on portfolio  $p$ ,  $R_{ft}$  is the riskfree rate of return,  $RMRF_t$  is the market return in excess of the riskfree rate,  $SMB_t$  is the difference between the value-weighted return of a portfolio of small stocks and the value-weighted return of a portfolio of large stocks,  $HML_t$  is the difference between the value-weighted return of a portfolio of high B/M stocks and the value-weighted return of a portfolio of low B/M stocks,  $UMD_t$  is the difference between the value-weighted return of a portfolio of stocks with high returns during months  $t - 12$  to  $t - 2$  and the value-weighted return of a portfolio of stocks with low returns during months  $t - 12$  to  $t - 2$ ,  $ISEN_t$  is the difference between the equal-weighted return of a portfolio of stocks with high buy-sell imbalance (top three deciles) and the equal-weighted return of a portfolio of stocks with low buy-sell imbalance (bottom three deciles) in month  $t$ , and  $\varepsilon_{pt}$  is the residual return on the portfolio. The individual investor holdings data are from a large discount brokerage house in the U.S. for the 1991-96 time-period. \* and \*\* denote significance at the 5% and 1% levels, respectively.

*Panel A: Mean Arbitrage Cost Estimates*

Sorting Variable	Quintile Portfolios				
	Low	Q2	Q3	Q4	High
<i>Firm Size</i>	331.49	158.67	122.14	92.99	63.68
<i>Book-To-Market</i>	359.15	224.43	179.14	170.61	221.54
<i>Insti. Own.</i>	364.01	248.38	183.51	137.79	102.79
<i>Stock Price</i>	589.44	290.25	164.85	115.89	85.12

*Panel B: Sentiment Factor Loadings*

Habitat	Arbitrage Cost Quintile Portfolios				
	Low	Q2	Q3	Q4	High
<i>Small Firm Size</i>	0.082(350)	0.173**(519)	0.539**(723)	0.545**(850)	0.580**(804)
<i>High Book-To-Market</i>	0.020*(179)	0.183**(298)	0.456**(339)	0.467**(267)	1.163**(168)
<i>Low Insti. Own.</i>	0.006(243)	0.201*(84)	0.671**(96)	0.298*(145)	0.586*(228)
<i>Low Stock Price</i>	-0.335(7)	-0.165(43)	0.428**(167)	0.483**(426)	0.543*(602)



**Figure 1. Residual buy-sell imbalance (RBSI) correlation distribution.** This figure shows the residual BSI correlation distribution obtained by forming 1,000 pairs of non-overlapping 250-stock random-portfolios. The BSI of portfolio  $p$  in month  $t$  is defined as,  $BSI_{pt} = \frac{100}{N_p} \sum_{i=1}^{N_p} BSI_{it}$  where the BSI for stock  $i$  in month  $t$  is defined as  $BSI_{it} = \frac{[\sum_{j=1}^{D_t} (VB_{ijt} - VS_{ijt})]}{[\sum_{j=1}^{D_t} (VB_{ijt} + VS_{ijt})]}$ . Here,  $D_t$  is the number of days in month  $t$ ,  $VB_{ijt}$  is the buy volume (measured in dollars) for stock  $i$  on day  $j$  in month  $t$ ,  $VS_{ijt}$  is the sell volume (measured in dollars) for stock  $i$  on day  $j$  in month  $t$ , and  $N_p$  is the number of stocks in the portfolio. Residual BSI is obtained by removing the common dependence of portfolio BSI on the market. The individual investor holdings data are from a large discount brokerage house in the U.S. for the 1991-96 time-period.