Retail Investor Sentiment and Return Comovements

ALOK KUMAR and CHARLES M.C. LEE*

ABSTRACT

Using a database of more than 1.85 million retail investor transactions over 1991 to 1996, we show that these trades are systematically correlated − i.e., individuals buy (or sell) stocks in concert. Moreover, consistent with noise trader models, we find that systematic retail trading explains return comovements for stocks with high retail concentration (i.e., small-cap, value, lower institutional ownership, and lower-priced stocks), especially if these stocks are also costly to arbitrage. Macroeconomic news and analyst earnings forecast revisions do not explain these results. Collectively, our findings support a role for investor sentiment in the formation of returns.

THIS PAPER USES THE TRADING RECORDS of individual investors to investigate the effect of retail trading on stock returns. Our analysis is based on a sample of more than 1.85 million buy and sell transactions made by over 60,000 retail clients at a large U.S. discount brokerage firm. In the first part of the study, we examine whether the buy-sell activities of retail investors contain a common directional component. In the second part of the study, we measure changes in investor sentiment based on the direction of these retail trades, and evaluate the impact of retail investor trading on comovement in stock returns. Specifically, we examine which stocks, in the cross-section, are most affected by systematic retail trading.

Our investigation is motivated by two alternative views of return comovements. The

---

* Alok Kumar is at the Mendoza College of Business, University of Notre Dame. Charles Lee is at Barclays Global Investors and the Johnson Graduate School of Management, Cornell University. We would like to thank an anonymous referee, an associate editor, 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, Terrance Odean, Ted O'Donoghue, Galina Ovtcharova, Vicente Pons, Paul Schultz, Mark Seasholes, Robert Stambaugh (the editor), Kent Womack, and Jeff Wurgler, as well as seminar participants at Cornell, the University of Illinois (Chicago), London Business School, Massachusetts Institute of Technology, Notre Dame, the 2003 WFA Meetings (Los Cabos, Mexico), and the Spring 2004 NBER Behavioral Finance Working Group Meetings (Chicago, IL) 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. The I/B/E/S data were provided by Thomson Financial as part of a broad academic program to encourage earnings expectations research. Finally, we thank Lubos Pástor for providing the liquidity factor data. All remaining errors and omissions are our own.

traditional view posits that the current price of a stock closely reflects the present value of its future cash flows. According to this view, the correlations in the returns of two assets arise from correlations in the changes in the assets' fundamental values, with demand shocks or shifts in investor sentiment playing no role because the actions of arbitrageurs readily offset such shocks.

An alternative theory argues that the dynamic interplay between noise traders and rational arbitrageurs establishes prices (e.g., Shiller (1984), Shleifer and Summers (1990)). According to this second view, in addition to innovations in fundamentals, factors such as the correlated trading activities of noise traders also induce comovements and arbitrage forces may not fully absorb these correlated demand shocks.[1]

Our analysis tests a particular form of the noise trader model in which individual (i.e., retail) investor sentiment can affect stock returns. We have in mind a clientele-based model that closely parallels the models of Bodurtha, Kim, and Lee (BKL; 1995) and Barberis, Shleifer, and Wurgler (BSW; 2005). In these models, different investor groups restrict themselves to trading within different natural "habitats," or groups of stocks. Thus, the returns of individual stocks reflect not only fundamental risk, but also changes in the systematic time-varying preferences (i.e., "sentiment") of important investor groups.[2]

Our main conjecture is that systematic trading by retail investors could lead to stock return comovements beyond the usual risk factors. The U.S. equity market is characterized by widespread direct stock ownership by retail investors. Extant evidence shows that these investors spend far less time on investment analysis, they engage in more attention-based trading, and they typically rely on a different set of information sources than their professional counterparts.[3] If the buy-sell patterns of retail investors do not move in lock-step with overall market movements, assets in market segments dominated by these investors could be characterized by pricing anomalies that are associated with their trading activities. Our tests explore this possibility.

In the first part of the study, we show that retail investors' trades are systematically correlated, that is, individuals tend to buy or sell stocks in concert with each other. Specifically, we document two related findings: (1) a strong positive correlation in the buy-sell imbalance (BSI) of retail investors across non overlapping portfolios of different stocks, that is, when retail investors buy (sell) one group of stocks, they tend to buy (sell) other groups of stocks; and, (2) correlated trading behavior holds across different individuals, that is, when one set of retail investors buys (sells) stocks, a different set of retail investors also tends to buy (sell) stocks.[4] These findings indicate the existence of a systematic (or common directional) component in the trading activities of retail investors.

In the second part of the paper, we examine whether the systematic component of retail trades, which we dub "retail investor sentiment," has incremental power in explaining return comovement. To measure changes in retail sentiment for a certain basket of stocks, we construct a buy-sell imbalance (BSI) measure for various stock portfolios. We then estimate multifactor time-series models in which we use the portfolio buy-sell imbalance as one of the explanatory variables.

For stocks with high retail concentrations, we find that a portfolio-level BSI measure has a significant incremental ability to explain return comovements. This result holds even after controlling for the effects of both innovations in macroeconomic variables (unexpected inflation, monthly growth in industrial production, change in term spread, and change in value spread) and empirically inspired risk factors, namely, the market excess return (RMRF), the size factor (SMB), the book-to-market (B/M) factor (HML), and the momentum factor (UMD). Although our BSI measure is correlated with changes in closed-end fund discounts (which Lee, Shleifer, and Thaler (1991) use to proxy for retail sentiment), it has much greater explanatory power for returns than the closed-end fund variable.

Consistent with clientele-based models of return comovement (e.g., BSW (2005)), we find that the sensitivity of firms to the BSI variable is a function of retail concentration, or retail trading intensity. Specifically, we find that small firms, lower priced firms, firms with lower institutional ownership, and value (high B/M) firms, all are associated with strong retail concentrations and disproportionately high retail trading activities. These same firms also all have positive and significant loadings on BSI. Among the stocks in the top quintile of these characteristics, a one-standard deviation change in monthly BSI corresponds to a 0.60% change in average monthly returns (controlling for all other factors).

Also, consistent with the prediction of noise trader models, we find that the importance of BSI in explaining returns is a function of arbitrage costs. Among firms with a high retail concentration, firms that are more difficult to arbitrage consistently exhibit larger BSI loadings.[5] In fact, we find that portfolio BSI loadings are a (nearly monotonic) function of arbitrage costs in all four measures of retail investor habitat (i.e., small, low price, high B/M, and low institutional concentration stocks). For firms with both high retail concentration and high arbitrage costs, a one standard-deviation increase (decrease) in BSI corresponds to an increase (decrease) of around 1.00% in average monthly returns.

Collectively, our findings are relevant to the debate on whether investor sentiment plays a role in financial markets. The traditional case against such a role for investor sentiment in markets is based on two key assertions: (1) the cognitive foibles that individuals commit 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, and thereby render prices unaffected.[6]

Our analysis speaks directly to these issues, suggesting that, at least in the case of retail investors, both of the above 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 results therefore support a friction- or sentiment-based theory of returns comovement, such as BSW (2005). BSW argue that the observed return patterns that obtain around the inclusion or deletion of a stock from the S&P500 stock index is clientele-related. Our evidence suggests that their habitat-based model applies not only to institutional indexers, but also to retail investors.

Note that the analysis here does not allow us to identify the precise driver of retail investor trading, that is, whether the time-varying preferences of retail investors are due to liquidity concerns, risk aversion, or irrational sentiment. However, we are able to eliminate some likely suspects. We show that the systematic component of retail sentiment is influenced by market-wide index returns, and innovations in a few macroeconomic factors. Controlling for these market-wide factors, as well as changes in analyst forecasts of future earnings, has little effect on our findings, however. Thus, the systematic component of retail sentiment does not appear to be a simple artifact of what we commonly perceive as fundamental news.

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 has proven to be controversial, however, spawning a number of follow-up studies.[7] A key point of contention is whether closed-end fund discounts are a good proxy for retail investor behavior. We shed light on this debate by providing direct evidence that links retail trades and stock returns. Consistent with LST, our results show that the directional trades of retail investors are associated with the excess returns of small firms and firms characterized by low institutional ownership. In addition, we show that for our sample period 1991 to 1996, the explanatory power for return comovements is significantly greater for BSI than for changes in closed-end fund discounts.

Finally, our paper is related to a growing literature in behavioral finance that examines the correlated trading behavior of retail investors and its impact on stock returns. Using a Chinese data set, Feng and Seasholes (2004) find that the trading activities of investors that live within a certain geographic region are strongly correlated. Similar in spirit, Jackson (2003) provides additional evidence of systematic trading patterns among Australian investors. Also, as we mention earlier, Barber, Odean, and Zhu (BOZ; 2003) provide evidence of correlated trading among retail investors in the U.S., and explore psychology-based explanations for these patterns. Our findings are consistent with these studies and we extend this line of inquiry by linking the correlated trading behavior of individual investors to stock returns. We believe that, at a minimum, our results highlight the need to study further the role of investor behavior in financial markets.[8]

The remainder of our paper is organized as follows. In the next section, we describe the data and our sentiment change measure and we document the existence of a directional component in the trading activities of retail investors. In Section II, we examine the sentiment-return relation for size-sorted portfolios, and perform a variety of robustness tests. We test specific predictions of the BSW model in Section III using habitat-sorted and arbitrage cost-sorted portfolios. Section IV concludes with a brief summary and discussion.

I. Evidence of Market-Wide Systematic Component

Our first set of tests is designed to evaluate the extent to which the trading activities of retail investors are correlated. An important assumption common to all noise trader models is that uninformed noise trader demand aggregates across a population of individuals, that is, individuals buy or sell baskets of stocks in concert with each other. In the absence of this type of systematic behavior, it is unlikely that noise trader sentiment can affect returns.[9]

To examine whether retail investors trade in concert, we measure the correlations in the buy-sell imbalance (BSI) time series of pairs of non overlapping stock portfolios. In addition, we examine the correlation in buy-sell behavior across individual investors (similar to BOZ (2003)). Finally, we conduct a series of variance-based tests with the same objective.

A. Data

The primary data for our study consist of trades and monthly portfolio positions of the retail investors at a major U.S. discount brokerage house over the period 1991 to 1996. While there are 77,995 households in the database, we focus on the 62,387 that trade 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 four-stock portfolio (median is three) 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% 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 nine trades per year. The average trade size is $8,779 (median is $5,239).[10]

Table I presents summary statistics for the trading activities of our sample investors. The individual investors in our sample execute 26,000 trades in a typical month or 1,244 trades on a typical day. As Panel A shows, the number of buy trades in any given year is higher than the number of sell trades. Furthermore, in a given year, the group of retail investors in our sample trades approximately 6,000 to 7,000 stocks, suggesting that their trades spans a large set of stocks. The stock and investor-level trading statistics (Panels B and C, respectively) show that in a typical month, approximately 20% stocks have five or more trades, and roughly 2% to 7% investors execute five or more trades.

******* Table I here *******

In addition to the retail investor data, we obtain quarterly institutional ownership data for our sample stocks from Thomson Financial. The data contain the quarter-end 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.[11] Using the quarterly institutional holdings, we compute the aggregate institutional ownership for each stock at the end of each quarter. We then use this aggregate ownership data to construct year-end institutional ownership portfolios.

We use several other standard data sets in this study. For each stock in the sample, we obtain monthly price, returns, and market capitalization data from Center for Research on Security Prices (CRSP) and quarterly book value of common equity data from COMPUSTAT. The monthly time series of the three Fama-French factors, the momentum factor, the NYSE size break-points, and the book-to-market break-points for each month come from Ken French's data library.[12] Finally, we obtain analysts' quarterly earnings estimates from the I/B/E/S summary files.

B. Measuring Changes in Retail Sentiment

We use the trading activities of retail investors to measure changes in their sentiments. Shiller (1984) suggests that common sentiments arise when investors trade on pseudo-signals such as price and volume patterns, popular forecasting models, or the forecasts of Wall Street gurus.[13] Prior evidence (e.g., Lee (1992), Odean (1999), Dhar and Kumar (2001), Barber and Odean (2003)) suggests that, indeed, at least a portion of the trading by retail investors is likely to be induced by pseudo-signals. Here, while we do not claim that systematic patterns in retail trades are necessarily non-fundamental in nature, to the extent that a systematic directional pattern is not explained by known risk factors, we refer to it as "retail sentiment."

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) over a particular time period t. To compute the monthly portfolio BSI, we first define the month-t BSI for stock i as:

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

Here, D_t is the number of days in month t, VB_{ijt} (VS_{ijt}) is the dollar-denominated buy (sell) volume for stock i on day j of month t.[14] A given month's stock-level BSI indicates whether, at an aggregate level, retail investors are net buyers (stock BSI > 0, that is, a positive change in their aggregate stock sentiment) or net sellers (stock BSI < 0, that is, a negative change in their aggregate stock sentiment) of a given stock over a given month. In other words, the stock-level BSI measure is a directional indicator of net retail demand for that stock in a given month.[15]

Next, we generate a measure of portfolio BSI by calculating an equal-weighted average of individual stock BSIs. That is, we solve:

$$BSI\,pt = \frac{100}{Np} \sum_{i=1}^{Np} BSI\,it\,. \tag{2}$$

Here, N_p is the number of stocks in portfolio p. This monthly portfolio-level BSI provides a measure of the number of investor buys minus sells for each stock, where each stock is weighted equally in the portfolio. Out intent is to capture the mean retail sentiment shift for stocks in the portfolio, with equal weight given to each stock. For example, consider a portfolio of the four stocks A, B, C, and D, which have prices of \$80, \$10, \$10, and \$10, respectively, in a particular month. Assume further that investors sell 100 shares of stock A, and buy 100 shares each of stocks B, C, and D in the specified month. Given this trading pattern, the BSI for stock A would be 1 while that of stocks B, C, and D would be −1. The portfolio BSI would be (−1 + 1 + 1 + 1)/4 = 0.50, indicating a bullish sentiment shift. In short, our portfolio BSI is a reflection of net investor demand across the stocks within the portfolio.

An alternative approach is to compute the aggregate dollar volume in-flow (*AVB*) and aggregate dollar volume out-flow (*AVS*) for all the stocks in a portfolio, defining the portfolio BSI as *(AVB−AVS)/(AVB+AVS)*.[16] However, under this alternative approach, a particular month's portfolio BSI can be strongly influenced by a single stock. This is problematic, especially in the case that a stock experiences unusually high trading volume due to an information event such as an earnings announcement or a stock recommendation change. Additionally, such a measure is sensitive to within-portfolio changes in the stock price distribution. Our equal-weighted measure avoids such biases.[17]

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

$$BSI_{pt} = b_0 + b_1\, RMRF_t + \varepsilon_{pt}. \tag{3}$$

Here, BSI_{pt} is the month-t buy-sell imbalance index for portfolio p, $RMRF_t$ is the month-t market return in excess of the risk-free rate, and ε_{pt} is the month-t residual BSI for portfolio p. The purpose of this regression is to remove the common component in investor net demand that is due to overall market movements. In all our empirical analysis, we use this orthogonalized measure of investor trading activity.[18]

C. Correlations among Random Non Overlapping Portfolios

To examine correlations among stock portfolios, we form 1,000 pairs of non overlapping stock portfolios that consist of k stocks (k = 50, 75, 100, 125, 250, and 500), where stocks are chosen randomly from the set of stocks traded by our sample investors. For each of these randomly chosen portfolios, we use equation (2) to construct a monthly BSI time series for the 71-month sample period from January 1991 to November 1996. We then orthogonalize these monthly BSI measures with respect to the market index return (equation (3)). Finally, we compute the correlations between pairs of BSI indices that we derive from non overlapping portfolios, and we generate an empirical distribution of BSI correlations.

Table II, Panel A reports the correlation statistics for different portfolio sizes and Figure 1 shows an empirical distribution of the pair-wise correlations for k = 250. The average BSI correlation is positive and significantly different from zero (p-value < 0.05) for all chosen portfolio sizes. We find that the average BSI correlation increases with portfolio size. For instance, for 250-stock portfolios, the average BSI correlation is 0.600; for 50-stock portfolios, this measure is 0.234. These results indicate the presence of a systematic component in the trading activities of our sample investors, a component that is uncorrelated with movements in the market index.[19]

******* Table II here *******

******* Figure 1 here *******

To obtain an estimate of the average BSI correlation in the absence of a systematic component in sample investors' trading activities (i.e., a benchmark average BSI correlation or expected BSI correlation), we also conduct Monte Carlo simulations. Specifically, we generate a BSI matrix, where, for each stock, we keep the frequency and timing of trades fixed but randomly assign a BSI in each month, with BSI lying uniformly in the range (−1, 1).[20] Using this simulated BSI matrix and following the procedure described earlier, we generate an empirical distribution of BSI correlations by computing the correlations between 1,000 pairs of non overlapping 250-stock portfolios.[21] We repeat this entire process 500 times and obtain a distribution of the average BSI correlation. We find that the average 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 BSI correlation is lower than 0.60 in each of our 500 repetitions. In short, the Monte Carlo results indicate that the observed BSI correlations are unlikely to be a chance occurrence (p-value < 0.01).

D. Correlations among Random Subsets of Investors

To further examine the nature of systematic trading behavior among retail investors, we explore the BSI correlations between pairs of non overlapping investor groups. Similar to our portfolio-based approach, we form 1,000 pairs of non overlapping investor groups, where each investor group contains k investors (k = 500, 1,000, 1,500, 2,000, 2,500, and 5,000) and investors are chosen randomly from the set of sample investors. Next, for each of these randomly chosen investor groups, we construct a monthly buy-sell imbalance (BSI) time series for the 71-month sample period of January 1991 to November 1996. The month-t BSI for investor group i is then:

$$BSI_{it} = \frac{\sum_{j=1}^{N_i} (VB_{jt} - VS_{jt})}{\sum_{j=1}^{N_i} (VB_{jt} + VS_{jt})}. \tag{4}$$

Here, N_i is the number of investors in group i, VB_{jt} is the month-t dollar-denominated buy volume measured across all stocks for investor j, and VS_{jt} is the month-t dollar-denominated sell volume measured across all stocks for investor j. As before, we orthogonalize the BSI measure relative to the movements of the market index. Finally, we compute the correlations between the pairs of BSI indices of these non overlapping investor groups and we generate an empirical distribution of BSI correlations.[22]

Table II, Panel B reports the correlation statistics for different sizes of investor groups. Analogous to our portfolio-based results, the average BSI correlation is positive and statistically significant (p-value < 0.05) for all selected investor group sizes. Furthermore, the average BSI correlation increases monotonically as group size increases. For instance, for investor groups that consist of 1,000 investors each, the average BSI correlation is 0.210, whereas for groups that consist of 5,000 investors, the average BSI correlation is 0.496. These results provide further evidence of systematic trading by retail investors in our sample.[23]

In sum, our correlation tests show that correlations between BSI indices are significantly positive over non overlapping stock portfolios as well as non overlapping investor groups. Additionally, our Monte Carlo tests suggest these strong positive correlations could not have occurred by chance.

E. An Alternative Test of Systematic Trading

We can also examine the existence of a systematic component in retail trading using an alternative test.[24] In the absence of a systematic component in retail trading, the BSI time series of an aggregate portfolio that contains all stocks in the sample will realize a significantly lower variance than the mean BSI variance of individual stocks. Particularly, in the extreme case in which investors execute unrelated random trades, the aggregate portfolio BSI variance should be statistically indistinguishable from zero.

Given this motivation, we compute a test statistic for the extent of correlated trading based on the ratio of the variance of the aggregate portfolio BSI and the mean BSI variance of individual securities (or k-stock portfolios). Unfortunately, due to sparse trading, firm-level BSI measures are noisy. Thus, we compute the variance for k-stock portfolios and explore the results for various measures of k. Furthermore, because we can construct a large number of k-stock portfolios, we adopt a procedure whereby all stocks in a given draw are assigned to k-stock portfolios on a random basis. We then repeat the process 500 times to obtain a distribution of the ratio of variances. If net buying demand is uncorrelated, our test statistic should approach zero as k approaches one.

The variance-based test results are consistent with our prior findings. When k = 100, the mean of the ratio of the aggregate portfolio variance and the mean k-stock portfolio variance is 0.619, which is reliably different from zero (p-value < 0.001). For k = 50, the mean variance ratio is 0.431; for k = 10, the test statistic is 0.226 (p-value < 0.001 for both). We also computed the test statistic (k = 1) for the 50 most actively traded stocks in our sample, and find it is equal to 0.128. In all instances, the variance ratio is reliably positive. In sum, the net demand of retail investors in our sample appears to be driven to a considerable degree by a common factor.[25]

In sum, our various test results consistently suggest the existence of a systematic (or common directional) component in the trades of retail investors. In particular, when retail investors buy (sell) one basket of stocks, they are likely to simultaneously buy (sell) other stock baskets. Similarly, when one set of investors buys (sells) stocks, other sets of individuals also tend to be buying (selling).

II. Retail Sentiment Shifts and Stock Returns

The existence of a common directional component in the trading activities of retail investors suggests that changes in retail sentiment might induce comovements in stock returns. In this section, we examine this possibility within a multi factor time-series modeling framework.

A. Multi Factor Time-Series Model

To examine the incremental ability of retail sentiment shifts 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 in which the first three factors are those of Fama and French (1992, 1993), the fourth factor is the momentum factor (e.g., Jegadeesh and Titman (1993), Carhart (1997)), and the fifth factor is the appropriate portfolio BSI measure. That is, we estimate the following factor model:

$$
\begin{aligned}
R_{pt} - R_{ft} = \; & \alpha_p + \beta_{1p}\, RMRF_t \\
& + \beta_{2p}\, SMB_t + \beta_{3p}\, HML_t + \beta_{4p}\, UMD_t \\
& + \beta_{5p}\, BSI_{pt} + \varepsilon_{pt}.
\end{aligned}
\tag{5}
$$

Here, R_{pt} is the portfolio rate of return, R_{ft} is the risk-free rate of return, $RMRF_t$ is the market return in excess of the risk-free 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 $t-2$, BSI_{pt} is the equal-weighted BSI of stocks in portfolio p, and ε_{pt} is the residual return on the portfolio.

In our asset pricing tests, we examine the incremental explanatory power of portfolio-level BSI measures (i.e, portfolio-level sentiment changes) rather than the market-wide BSI measure (i.e, aggregate sentiment changes). As expected, portfolio-level BSI measures are highly correlated with the market-wide BSI measure. For example, the correlations between market-wide BSI and the BSI measures for individual size quintile portfolios range from 0.714 to 0.890. However, we use portfolio-level BSI measures because we find that the market-wide measure is not a sufficient statistic for individual portfolio BSI measures.

In ancillary tests, we find that the mean BSI correlations for non overlapping portfolios are higher when stocks are selected from the same stock category than when stocks are selected from different categories. In other words, within-category correlations are reliably higher than cross-category correlations. This pattern obtains across stock categories defined using size, B/M, price, and institutional ownership.[26] Moreover, we find that the BSI correlations *across* stock categories are reliably lower than unity, suggesting that a market-wide measure is likely to omit some information that is contained in portfolio-level BSI measures.[27] For all these reasons, we use portfolio-level BSI measures computed for each stock category, rather than a market aggregate BSI measure, in the following tests.

B. Multi Factor Model Estimation Results for Size-Sorted Portfolios

To set the stage, we consider portfolios obtained by sorting on firm size. At the end of each year, we sort the entire universe of stocks for which returns data are 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 does not change during the course of the year. Portfolio 1 consists of stocks with the lowest market capitalization, while portfolio 5 contains stocks with the highest market capitalization. Next, for each portfolio, we compute the monthly portfolio return as an equal-weighted average of all stocks in the portfolio and we construct a monthly portfolio return time series. Finally, we estimate the time-series factor model (equation (5)) for each size quintile portfolio.

Table III reports the relation between our measures of retail sentiment shifts and four empirically inspired risk factors that are common in the literature. Panel A presents descriptive statistics for the BSI time series in each size quintile portfolio. We find that the BSI time-series for the small-cap portfolio realizes higher volatility. Furthermore, Panel B shows that the portfolio BSI measures are moderately correlated with the four risk factors (market or RMRF, small-minus-big or SMB, high-minus-low or HML, and momentum or UMD). In subsequent tests, we control for the effects of these standard risk factors when estimating the relation between retail sentiment shifts and stock returns.

******* Table III here *******

Table IV presents the time-series factor model estimates for each of the five size-quintile portfolios. For the small-cap (size quintile 1) stock portfolio, the BSI loading is positive (0.069) and statistically significant (*t*-statistic = 3.030), whereas for the remaining four size quintile portfolios, the portfolio BSI factor loadings are small in magnitude and statistically insignificant. For small-cap (Q1) stocks, α and the loading on *UMD* both become insignificant with the inclusion of BSI, suggesting that retail sentiment helps explain small-firm excess returns through an interaction with the momentum factor.

******* Table IV here *******

Collectively, our findings indicate that retail sentiment shifts have incremental ability to explain return comovements among small-cap stocks. Consistent with behavioral theory, stocks in the lowest size quintile earn positive (negative) excess returns when retail investor sentiment grows more bullish (bearish). However, we find no significant relation between retail sentiment shifts and the returns of other size quintile portfolios.[28] Later in the paper (see Section III.D), we conduct additional tests to evaluate both the economic significance of BSI in explaining small-cap stock returns, and the incremental effect of arbitrage cost.

C. Do Retail Sentiment Shifts Reflect Innovations in Macroeconomic Variables?

It is possible that BSI reflects the response of retail investors to innovations in macroeconomic variables such as those documented in Chen, Roll, and Ross (CRR; 1986). To examine this possibility, we follow CRR and consider the following four macroeconomic variables: (i) *UI*: unexpected inflation, where we use the average of the 12 most recent inflation realizations to estimate the expected level of inflation, (ii) *MP*: monthly growth in industrial production, (iii) ΔTS: 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 three-month Treasury bill, and (iv) ΔRP: change in the risk premium, where the risk premium represents the difference between the yields of Moody's BAA-rated and AAA-rated corporate bonds. Additionally, we consider innovations in monthly unemployment rate ($\Delta UNEMP$) and innovations in average hourly earnings ($\Delta WAGE$) as potential determinants of BSI.

In unreported analysis, we find that the portfolio BSI measures are moderately correlated with innovations in macroeconomic variables – in absolute terms, the correlations lie in the range of 0.02 to 0.40.[29] The contemporaneous correlations with *UI* and ΔRP are negative, while the correlations with *MP* and ΔTS are mostly positive. The correlations with $\Delta UNEMP$ and $\Delta WAGE$ have mixed signs but they are small in absolute terms (range is 0.02 to 0.07). This evidence suggests that changes in retail investor sentiment are likely to be influenced to some extent by innovations in macroeconomic variables. Nevertheless, they are unlikely to be fully explained by these innovations.

To evaluate the incremental power of BSI for explaining return comovement in small-cap stocks, we remove the common dependence of our portfolio BSI measures on innovations in macroeconomic variables. Specifically, we perform the following regression for each portfolio-level BSI time series:

$$BSI_{pt} = b_0 + b_1 MACRO_t + \varepsilon_{pt}. \tag{6}$$

Here, BSI_{pt} is the month-t buy-sell imbalance measure for portfolio p, $MACRO_t$ is a vector of month-t innovations in macroeconomic variables (i.e., *UI, MP,* ΔTS, ΔRP, $\Delta UNEMP$, and

$\Delta WAGE$), and ε_{pt} is the month-t residual BSI for portfolio p.[30]

Using the residual portfolio BSI measure, we again estimate the time-series factor model (equation (5)) for the quintile 1 size portfolio. Our estimation results indicate that the BSI loading remains virtually unchanged. The BSI loading estimate for the portfolio (see Table IV) is 0.064 with a t-statistic of 2.678 (only marginally different from our previous estimate of 0.069 with a t-statistic of 3.030 reported in Table IV). This result suggests that even though changes in retail sentiment are likely to be partially induced by innovations in macroeconomic variables, the residual retail sentiment changes as captured by the residual BSI measure have considerable incremental power for explaining comovements in small-cap stock returns.[31]

D. Do Retail Sentiment Shifts Reflect Innovations in Stock Fundamentals?

In addition to innovations in macroeconomic variables, the BSI may be influenced by changes in investors' expectations about future cash flows. We use revisions in analysts' forecasts of future earnings as a proxy for changes in investors' expectations about future cash flows. We add the mean change in analysts' forecasts of quarterly earnings (EFC) to the time-series model estimated above (equation (6)). Specifically, we perform the following regression for each portfolio-level BSI time series:

$$BSI_{pt} = b_0 + b_1 MACRO_t + b_2 EFC_{pt} + \varepsilon_{pt}. \qquad (7)$$

Here, EFC_{pt} is the mean change in the analysts' month-t forecasts of quarterly earnings for stocks in portfolio p. Other variables in the model are as defined earlier.

Examining the correlations between EFC and portfolio BSI measures, we find that the portfolio BSI measures are weakly correlated with earnings forecast changes. The correlations vary between −0.04 and 0.06 for size quintile portfolios and between −0.18 and 0.05 for B/M quintile portfolios. For other portfolios, the correlations are less than 0.20 in absolute terms.

When we use the residual portfolio BSI measure and estimate the time-series factor model (equation (5)) for the quintile 1 size portfolio again, we find that the BSI loading estimate is 0.064 (t-statistic = 2.601). This estimate is very similar to the BSI loading that we obtain when we only control for the effects of innovations in macroeconomic variables. This evidence indicates that portfolio BSI is not significantly influenced by changes in investors' expectations about future cash flows. This result, in combination with our earlier findings, suggests that portfolio BSI is not simply an artifact of the news events usually associated with changes in stock fundamentals.

E. Portfolio BSI and Closed-End Fund Discount Changes

Previous studies (e.g., Lee, Shleifer, and Thaler (1991)) employ changes in closed-end fund discount rates as a proxy for retail investor sentiment shifts. In this subsection, we examine the empirical relation between BSI and changes in closed-end fund discounts, and we evaluate the incremental explanatory power of each variable for the returns of small stocks.

To conduct these tests, we first compute a value-weighted index of closed-end fund discount rates (*VWD*) following the methodology in Lee, Shleifer, and Thaler (1991). We then use monthly changes in this index (ΔVWD) as a proxy for retail investor sentiment shifts. Our tests evaluate the relative abilities of portfolio BSI and ΔVWD to explain small stock (size quintile 1) portfolio returns by estimating different specifications of the time-series factor model (equation (5)). In these tests, we use a factor model that includes the three Fama-French factors, the momentum factor, a portfolio-level BSI measure, and ΔVWD.

We find that the portfolio BSI variable is positively correlated (correlation = 0.156) with the ΔVWD variable, suggesting that when retail investors are buying, closed-end fund discounts tend to narrow.[32] Furthermore, when both variables are included in the factor model, we find that the portfolio BSI variable has a positive coefficient estimate (0.070 with a *t*-statistic of 3.223) while the coefficient estimate of the ΔVWD variable is statistically insignificant. These results suggest that our direct measure of changes in retail sentiment has a greater ability to explain comovements in small-cap stock returns than changes in closed-end fund discount rates.[33]

F. Do Retail Sentiment Shifts Reflect Correlated Informed Trading?

To better understand the nature of the retail sentiment measure, we also examined the possibility that BSI reflects correlated informed trading, that is, that on average retail investor trades are profitable. Specifically, we employ a methodology introduced in Odean (1999) to examine the profitability of retail trades. If these retail investors are collectively informed about the stocks they choose to trade, on average, the stocks they buy should outperform the stocks they sell. Consequently, the mean post-trade buy-sell return differential for a stock category provides a measure of the degree of informed retail trading within a stock category.

Our untabulated results show that the mean *k*-day post-trade return differential between retail buys and retail sales is negative for all five size categories, particularly for size quintile 1, for which the influence of retail trading is strongest. For instance, the mean one-year return (i.e., k = 252) following investor trades is −3.83% for the small-stock category and −2.96% for the large-stock category.[34] While this evidence suggests that investors are not responding to common information signals, it is possible that the mean return differential is negative because investors buy less risky stocks and sell more risky stocks. However, when we explicitly control for risk, the return differentials are even stronger, with an annual risk-adjusted return differential of −5.33%.[35] In sum, this evidence suggests that retail sentiment shifts are unlikely to reflect correlated trading induced by common information signals about certain subset of stocks.

G. Additional Robustness Tests

We also conduct a number of additional robustness tests to verify the sensitivity of our results to various model perturbations. In the first set of tests, we examine the effect of sparse trading on the incremental explanatory power of BSI for small-firm stock returns. Specifically, when computing the month-*t* BSI for portfolio *p*, we exclude all stocks that have fewer than *k* trades during the month. This test is motivated by the concern that the BSI measure might become quite noisy when there are only a few retail trades for a particular stock in a given month.

We find that the BSI loading estimates remain statistically significant for a fairly wide range of *k* values. For instance, when $k = 4$, the incremental BSI loading estimate for the size quintile 1 size portfolio, controlling for the other risk factors (see Table IV), is 0.054 with a *t*-statistic of 2.423. As *k* increases, the BSI loading estimate decreases, and it becomes statistically insignificant when $k = 8$. This is not surprising because when we employ the "minimum eight trades" filter, a large number of stocks (about 91% of the stocks in the original sample) are excluded from the BSI computation.

Given the sensitivity of our BSI loading estimates to a minimum number of trades cutoff, it is also likely that the BSI loading estimates are sensitive to volume turnover. Put differently, the existence of sparsely traded stocks may influence our main results. In our second set of tests, we examine the impact of trading volume on the strength of sentiment-return relation. Specifically, we estimate the BSI loading for double-sorted portfolios on firm size and volume turnover, where, as before, we carry out the two sorts independently. Within the small-cap stock category (size quintile 1), we find that the BSI loading estimate is *higher* for the portfolio with the highest (quintile 5) turnover. The BSI estimates (*t*-statistics) for turnover quintile portfolios 1-5 are 0.045 (3.628), 0.042 (2.618), 0.016 (0.567), 0.037 (1.307), and 0.160 (2.899), respectively. These results suggest that our BSI loading estimates are robust to concerns about sparse trading.

In a third set of tests, we examine the robustness of our findings to the use of a stock-level BSI measure. This test is motivated by the concern that our results might be related to the way buy and sell volume is aggregated for each stock in computing the portfolio BSI variable.[36] For this test, we compute a BSI measure for each stock and allow it to assume only one of three possible values: if the number of buys exceeds the number of sells, we set BSI equal to one; if the number of sells exceeds the number of buys, we set BSI equal to negative one; and if the number of buys and sells is equal, we set BSI equal to zero. The portfolio BSI is then simply the average of the individual firm-level BSIs.

We again estimate the time-series factor models using this modified portfolio BSI measure and find that all our results remain virtually unchanged. For instance, the BSI loading estimate for the quintile 1 size portfolio (see Table IV) is 0.067 with a *t*-statistic of 3.131 (only marginally different from our previous estimate of 0.070 with a *t*-statistic of 3.030) when the BSI measure is used as an explanatory variable along with the other four risk factors. Collectively, these results indicate that our BSI loading estimates are quite robust to perturbations in the construction of this variable.

III. The Effect of Sentiment in Other Retail "Habitats"

The BSW model of return comovements predicts that the sentiment-return relation is a function of investor habitat, that is, the sentiment-return relation is likely to be stronger among stocks that have a higher concentration of retail investors. To test this prediction, we examine whether the BSI loading estimates are stronger for stock categories that constitute the natural habitat of retail investors.

A. Identifying Retail "Habitat"

First, we 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 V reports the results of these tests. In Panel A, we report the mean percentage excess 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.

******* Table V here *******

Panel A results show that retail investors in our sample tend to concentrate their holdings in small, high B/M, low institutional ownership, 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, on average, retail investors exhibit a slight 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 which stocks retail investors own, 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 our sample retail investors, we obtain a normalized measure of retail trading activity (NTA) for each stock:

$$NTAit = \frac{\begin{array}{c}\text{Number of shares traded by}\\ \text{investors in our sample}\end{array}}{\begin{array}{c}\text{Number of shares traded}\\ \text{in the market}\end{array}}\ x\ 10^6, \qquad (8)$$

where NTA_{it} is the month-*t* normalized retail trading activity for stock. 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, 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 results show that the mean *NTA* decreases along size, institutional ownership, and stock price quintile portfolios, and 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 institutional-ownership, and high B/M stocks.

B. Retail Sentiment-Induced Comovements in Habitat-Based Portfolios

To examine the variation in the degree of return comovements generated by retail sentiment shifts (i.e., portfolio BSI), we perform one-dimensional sorts along the B/M, institutional ownership (IO), and stock price dimensions. The concentration of retail investors varies within each of these three stock categories. We carry out the five-factor model estimation (see equation (5)) 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 are available from CRSP according to their B/M ratio at the end of November 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 VI, Panel A.

******* Table VI here *******

We estimate the time-series factor model for each B/M quintile portfolio. Table VI, Panel B summarizes the estimation results.[37] We find that the BSI loading is positive and statistically significant for high B/M stock portfolios (B/M quintile portfolios 4 and 5). For B/M quintile portfolio 4, the BSI loading estimate is 0.020 (p-value < 0.05), while for the value stock portfolio (B/M quintile portfolio 5), the BSI loading estimate is considerably higher (0.060 with *p*-value < 0.01). 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 the retail concentration is higher. More specifically, the stocks in B/M quintile portfolios 4 and 5 earn positive (negative) excess returns when retail investor sentiment grows more bullish (bearish).

We also estimate the time-series factor model for each IO and price quintile portfolio. These results are also summarized in Table VI, Panel B. As expected, the BSI 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 BSI loading estimates are 0.071 (*p*-value < 0.01) and 0.069 (*p*-value < 0.01), respectively. The loadings decrease in magnitude and become statistically insignificant for 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 broadly consistent with the habitat view of comovement (Barberis, Shleifer, and Wurgler (2005)), which suggests that the impact of retail investor sentiment is stronger among stocks for which retail concentration is higher. Specifically, we find that when retail investors become more bullish (bearish), smaller stocks, value stocks, low IO stocks, and lower priced stocks earn positive (negative) excess returns. These results are also consistent with Daniel, Hirshleifer, and Subrahmanyam (2001), in whose model systematic factor loadings such as *SMB* and *HML* serve as proxies for mispricing.[38] To To the extent that the BSI loading estimates are significantly positive for small-cap (i.e., high *SMB*) and value (i.e., high *HML*) stocks, our findings support their prediction. Later, in the context of an analysis of arbitrage costs, we explicitly examine the effect of sorting by SMB and HML factor loadings.

C. Sentiment-Return Lead-Lag Relation

Our results so far are only able to establish a contemporaneous relation between portfolio BSI and portfolio returns. This evidence does not necessarily imply that correlated trading influences returns. To examine the sentiment-return dynamics more accurately, we focus on the lead-lag relation between portfolio BSI and portfolio returns. Using a bivariate vector autoregression (VAR) framework, we test the null hypothesis that portfolio BSI responds to stock returns but that the portfolio BSI measure itself has no ability to predict stock returns. The results are presented in Table VI, Panel C.[39] We find that *p*-values are less than 0.05 for small-cap, high B/M, low institutional ownership, and low stock price portfolios. In addition, other related portfolios have lower *p*-values. Overall, this evidence indicates that lagged portfolio BSI can predict portfolio returns for certain extreme portfolios in which the contemporaneous sentiment-return is strong, that is, portfolio BSI loading estimates are positive and significant (see Panel B). Consequently, the null of no return predictability can be rejected.

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 the time-series factor model (equation (5)) for four sets of portfolios formed by performing independent double sorts on arbitrage cost and size, B/M, IO, and stock price variables.

As a starting point, we follow Wurgler and Zhurvaskaya (2002) and use the variance of the residual from a capital asset pricing model (CAPM) regression (i.e., the idiosyncratic risk of each firm) as a proxy for arbitrage cost. Specifically, we use the monthly stock returns from the previous 60 months to estimate the CAPM regression. In later tests, we explore the robustness of our results to the use of other arbitrage cost measures (e.g., liquidity beta, HML and SMB factor loadings).

We present our main results in Table VII. Panel A reports the mean arbitrage cost estimates 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.

******* Table VII here *******

Of greater interest is the cross-sectional variation in the BSI loadings when firms are sorted on arbitrage cost. Panel B of Table VII reports the portfolio BSI loadings 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 BSI loading estimates.

We find that arbitrage costs have a strong influence on BSI loadings; specifically, the impact of retail sentiment is confined to high arbitrage cost portfolios. This result is robust across all measures of retail habitat. For instance, within the lowest size quintile portfolio (row 1), the BSI loadings increase across arbitrage cost quintile portfolios and is statistically significant for the top two quintiles (Q4 and Q5). When arbitrage costs are low (Q1 to Q3), the BSI loading is insignificant. We observe a similar pattern across the other three rows. In each instance, when arbitrage costs are higher, the loading on BSI is stronger.

Although idiosyncratic risk is a plausible arbitrage cost proxy, it might also proxy for other related constructs.[40] For robustness, we also experiment with another arbitrage cost proxy. Under the assumption that arbitrage costs are higher for less liquid stocks, we use the Pástor and Stambaugh (2003) liquidity beta as an arbitrage cost proxy.[41] Table VII, Panel C reports the estimates. Consistent with Panel B, we find that the BSI loadings are stronger for firms with lower liquidity, and that the pattern is robust across all four measures of retail habitat. In fact, the results are generally stronger using liquidity beta as an arbitrage cost proxy. In Panel C, BSI loadings are significant for most portfolios, including three of the four intermediate portfolios. Again, these results indicate stocks with higher arbitrage costs are more sensitive to retail sentiment.

Table VIII offers a more complete picture of the interaction between retail concentration and arbitrage costs. Table values represent BSI loadings for double-sorted portfolios formed using firm characteristics and arbitrage costs. These findings provide additional insights for intermediate portfolios when arbitrage costs and retail concentrations are not as high. In general, we find that sentiment shifts induce return comovements only when arbitrage costs are relatively high (Q4 or Q5 portfolios). Interestingly, in a few portfolios with low retail concentrations, the loading on BSI is reliably negative. This evidence suggests that retail investors generally serve as liquidity suppliers among these stocks, that is, their net demand for stocks is contrary to overall market movements. This is consistent with Kaniel, Saar, and Titman (2004), who also find that retail investors tend to provide liquidity when institutional buying (selling) pressure pushes prices up (down).[42]

******* Table VIII here *******

As a final robustness test, we also examine whether the BSI loading increases with the level of factor mispricing (as captured by the magnitude of *SMB* and *HML* loadings) within retail habitat portfolios. This test is motivated by the Daniel, Hirshleifer, and Subrahmanyam (2001) model, in which systematic factor loadings such as *SMB* and *HML* act as proxies for mispricing that result from limited arbitrage. Focusing our attention on quintile 1 size portfolios (to keep the concentration of retail activity constant), we find that BSI loading estimates do indeed increase with the level of mispricing. For instance, the BSI loading estimates (*t*-statistics) for *SMB*-sorted quintile portfolios are 0.019 (0.398), 0.028 (0.769), 0.106 (2.453), 0.026 (2.451), and 0.166 (3.458). Additionally, the BSI loadings increase monotonically across *HML* sorted portfolios. For *HML*-sorted portfolios, the BSI loading estimates (*t*-statistics) are 0.043 (0.731), 0.070 (1.521), 0.072 (1.803), 0.086 (1.229), and 0.143 (2.114). Consistent with the DHS model, the results indicate that the sensitivity to retail sentiment is higher when the factor mispricing (or alternatively, arbitrage cost) is higher.

Overall, consistent with sentiment-based models of return comovement, our results suggest that arbitrage cost and retail investor habitat are joint determinants of the strength of the sentiment-return relation. The retail sentiment shifts induce the greatest degree of comovement in the returns of stocks that belong to the retail habitat and are more difficult to arbitrage.

E. Economic Significance of Portfolio BSI Loading Estimates

Our results so far indicate a strong statistical relation between retail sentiment shifts and portfolio returns for certain subsets of stocks. But are these effects economically significant? To examine this question, we compute the change in the monthly portfolio-level expected return that corresponds to a one-standard deviation shift in the portfolio-level BSI measure. Because our focus is on the incremental effects of retail sentiment shifts, we first compute a residual BSI measure, using the following regression model:

$$BSI_{pt} = \alpha_p + \beta_{1p}\, RMRF_t + \beta_{2p}\, SMB_t + \beta_{3p}\, HML_t + \beta_{4p}\, UMD_t + \beta_{5p}\, MACRO_t + \varepsilon_{pt}. \qquad (9)$$

Here, BSI_{pt} is the month-t buy-sell imbalance index for portfolio p, *RMRF, SMB, HML,* and *UMD* are the four commonly used risk factors in month t, $MACRO_t$ is a vector of month-t innovations in macroeconomic variables (i.e., *UI, MP,* ΔTS, ΔRP, $\Delta UNEMP$, and $\Delta WAGE$), and ε_{pt} is the month-t residual BSI for portfolio p. A unit shift in the residual BSI corresponds to a shift in retail sentiment that is unaffected by the four risk factors or by innovations in macroeconomic variables.

Figure 2 reports measures of economic significance for portfolios in which retail sentiment shifts generate statistically significant comovements in returns. We find that for the small-firm portfolio, a one-standard deviation shift in the residual BSI (or retail sentiment) corresponds to a 0.56% monthly shift in the portfolio return. For the other three habitat measures, the corresponding shift in monthly average returns ranges from 0.52 (for value stocks) to 0.92% (for low IO stocks). These estimates indicate that moderate shifts in retail sentiment correspond to economically significant changes in portfolio returns.

******* Figure 2 here *******

This figure also shows the incremental impact of arbitrage costs. In all four retail habitats, higher arbitrage costs are associated with greater sensitivity to BSI. For instance, a one-standard deviation shift in residual BSI for small firms with high arbitrage costs is associated with a 0.96% shift in monthly returns. This effect is larger for the other three retail habitat proxies. Overall, Figure 2 results show that among stocks with a significant retail following, changes in retail sentiment are related to economically important changes in monthly returns.

IV. Summary and Conclusions

In this study, using a large data set of retail trades from a major discount brokerage house in the U.S., we examine the effect of retail trading patterns on comovement in stock returns. First, we document that the trading activities of retail investors contain a common directional component ─ when retail investors buy (sell) one group of stocks, they tend to buy (sell) other groups. Similarly, when some investors are buying (selling) stocks, other individuals also tend to be buying (selling). This evidence suggests that changes in portfolio-level retail sentiment may induce comovement in stock returns.

Next, using retail investors' trading activities, we obtain direct measures of retail investor sentiment changes and find that these measures have incremental explanatory power (over the standard risk factors and innovations in macroeconomic variables) 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 (bearish), the stocks in these portfolios enjoy higher (lower) excess returns.

Finally, we show that the strength of the sentiment-return relation 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, liquidity betas, etc.) exhibit much stronger sensitivity to changes in retail sentiment.

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 (2005). Consistent with the "habitat" version of the BSW model, we find that stocks preferred by retail investors are the ones most sensitive to shifts in retail investor sentiment. Also consistent with the predictions of their model, we find that the strength of the sentiment-return relation is a function of arbitrage costs.

More broadly, our results support a role for investor sentiment in the study of financial markets. The traditional case against such a role for investor sentiment 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. For instance, it would be interesting to examine whether the time-varying preferences of retail investors are due to liquidity concerns, risk aversion, or irrational sentiment. Our analysis suggests that retail sentiment is not a simple artifact of the news events that are generally associated with changes in stock fundamentals (e.g., macroeconomic news or analyst earnings forecast revisions). Nevertheless, questions remain as to the drivers of retail demand for stocks. 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 ongoing research.

REFERENCES

Baker, Malcolm, and Jeffrey Wurgler, 2005, Investor sentiment and the cross-section of stock returns, *Journal of Finance,* Forthcoming.

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, 2003, 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.

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

Barberis, Nicholas R., and Richard H. Thaler, 2003, A survey of behavioral finance, in George M. Constantinides, Milton Harris, and Rene Stulz, eds. *Handbook of Economics and Finance,* Chapter 18 (Elsevier Science B.V.).

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283-317.

Battalio, Robert H., and Richard R. Mendenhall, 2005, Earnings expectations, investor trade size, and anomalous returns around earnings announcements, *Journal of Financial Economics,* Forthcoming.

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, 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.

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

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.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 2001, Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* 56, 921-965.

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.

Dhar, Ravi, and Alok Kumar, 2001, A non-random walk down the main street: Impact of price trends on trading decisions of individual investors, Working paper, Cornell University.

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, 2004, Correlated trading and location, *Journal of Finance* 59, 2117-2144.

Frieder, Laura, and Avanidhar Subrahmanyam, 2005, Brand perceptions and the market for common stock, *Journal of Financial and Quantitative Analysis* 40*,* 57-85.

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.

Gompers, Paul A., and Andrew Metrick, 2001, Institutional investors and equity prices, *Quarterly Journal of Economics* 116, 229-259.

Hirshleifer, David, 2001, Investor psychology and asset prices, *Journal of Finance* 56, 1533-1597.

Hong, Harrison, Jeffrey Kubik, and Jeremy Stein, 2004, The only game in town: Stock-price consequences of local bias, Working paper, Princeton University.

Jackson, Andrew, 2003, The aggregate behavior of individual investors, Working paper, London Business School.

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

Jiang, Guohua, Charles M.C. Lee, and Yi Zhang, 2005, Information uncertainty and expected returns, *Review of Accounting Studies,* Forthcoming.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2004, Individual investor sentiment and stock returns, Working paper, Duke University.

Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, 1992, The impact of institutional trading on stock prices, *Journal of Financial Economics* 32*,* 23-43.

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

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., Andrei Shleifer, and Richard H. Thaler, 1991, Investor sentiment and the closed-end fund puzzle, *Journal of Finance* 46, 75-108.

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

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

Qiu, Lily, and Ivo Welch, 2004, Investor sentiment measures, Working paper, Brown University.

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, and Lawrence H. Summers, 1990, The noise trader approach to finance, *Journal of Economic Perspectives* 4, 19-33.

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.

Zhang, X. Frank, 2005, Information uncertainty and stock returns, *Journal of Finance*, Forthcoming.

Notes

[1] See Shleifer and Vishny (1997) for a theoretical exposition of this argument. Hirshleifer (2001) and Barberis and Thaler (2003) provide excellent surveys of the literature.
[2] 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.
[3] 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 (2005) discuss individual investor decision styles and information sources. Odean (1999) and Barber and Odean (2000, 2003) show that retail investors trade too often, particularly around news events. Finally, Lee (1992), Bhattacharya (2001), and Battalio and Mendenhall (2005) show that the response of small traders to earnings news differs sharply from those of large traders.
[4] This second finding is also reported by Barber, Odean, and Zhu (2003) [BOZ]. However, as we discuss later, their study explores factors that affect the degree of correlation across traders, and does not examine pricing implications.
[5] For most of our tests, we use the idiosyncratic risk of a stock as a measure of arbitrage cost (see Wurgler and Zhurvaskava (2002)). However, as we explain in detail later, we also find similar results with a liquidity beta measure (Pastor and Stambaugh (2003)) and factor loadings for the HML and SMB factors.
[6] See Shiller (1984), Shleifer (2000), or Lee (2001) for a more complete treatment.
[7] See Chen, Kan, and Miller (1993a, 1993b), Chopra, Lee, Shleifer, and Thaler (1993a, 1993b), Elton, Gruber, and Busse (1998), Gemmill and Thomas (2002), and Qiu and Welch (2004).
[8] Two other recent working papers explore the implications of correlated trading for stock returns. Hong, Kubik, and Stein (2004) use U.S. Census data to examine the asset pricing implications of investors' local biases. Their main empirical finding is that firms that are disproportionately large relative to regional economies trade at lower valuation multiples. Kaniel, Saar, and Titman (2004) use NYSE data to examine the role of individual trading as a predictor of short-horizon market-wide returns. We relate our findings to their study in more detail below (see Section III.D and footnote 40). Neither study examines the role of retail trades in explaining return comovements.
[9] For example, see the discussion in Chapter 1 of Shleifer (2000).
[10] See Barber and Odean (2000) for further details on the retail investor database.
[11] See Gompers and Metrick (2001) for further details on the institutional investor database.
[12] The data library is available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/`.
[13] Pseudo-signals refer to signals that are non informative in estimating a firm's fundamental value, but that may nevertheless be persuasive in their own right.
[14] The results are virtually identical when we use the number of shares instead of dollar volume to compute the monthly buy-sell imbalances.
[15] Given relatively sparse trading by our sample investors, a daily buy-sell imbalance (BSI) measure for a particular stock can be quite noisy. Therefore, for most of our empirical analyses, we use monthly aggregated measures of investor trades.
[16] The dollar volume-based BSI measure is identical to the D-ratio measure employed in the Lakonishok, Shleifer, and Vishny (1992) study to measure the excess demand of a group of institutional investors.
[17] Empirically, the correlation between aggregate equal-weighted BSI time series and value-weighted BSI time series is 0.590. As we show later, evidence of systematic correlated trading holds, although is weaker, when we use value-weighted rather than equal-weighted BSI measures.
[18] Consistent with other studies (e.g., Kaniel, Saar, and Titman (2004) and Barber and Odean (2000)), we find that retail investors are contrarians and tend to be net suppliers of liquidity to the market. In fact, the market-wide BSI measure has a correlation of –0.312 with market excess returns. The correlations between size-quintile portfolio BSIs and the market (RMRF) are reported in Table III, Panel B.
[19] These correlations remain significantly positive, but weaker, when we use value-weighted rather than equal-weighted portfolios of firm-level BSI. Specifically, for k-stock portfolios where k = 50, 75, 100, 125, 250, and 500, the mean correlations are 0.129, 0.173, 0.204, 0.244, 0.361, and 0.505, respectively.

[20] For robustness, we also conduct a similar test in which, instead of choosing BSI randomly from (−1, 1), we re-sample BSI from the observed empirical distribution of monthly BSI for each firm. In other words, we select the monthly BSI randomly with replacement from a firm's population of monthly BSI measures. The results are qualitatively identical to the findings that obtain using a random draw from a uniform distribution.

[21] Other choices for the portfolio size (e.g., 50, 100, and 500 stocks) produce similar results.

[22] Note that unlike our portfolio-level BSI, which is an equal-weighted measure, our investor-level BSI is a value-weighted measure. In other words, we wish to capture the dollar-weighted (rather than per capita) net buying demand for each investor group in a given month.

[23] Barber, Odean, and Zhu (2003) document similar results using a different methodology. Instead of choosing random subsets of investors, they divide the entire sample of retail investors into two groups such that group membership is assigned randomly. They then examine the correlation between the trading imbalances of the two non overlapping investor groups.

[24] We thank an anonymous referee for suggesting this test.

[25] Principal component analysis offers further insight on the presence of a systematic component in the time series of net buying by retail investors. Employing this methodology, we find that when k = 50, the first two principal components explain more than 50% (44.55% and 6.55%, respectively) of the variance in the portfolio BSI time series.

[26] For example, when both portfolios are chosen from size quintile 1, the mean 50-stock portfolio BSI correlation is 0.251. When one of the portfolio pairs is chosen from size quintile 1 and the other portfolio is chosen from size quintile 2 to 5, the mean 50-stock portfolio BSI correlations drops to 0.229, 0.209, 0.175, and 0.129, respectively. In other words, we find reliably lower cross-category correlations with quintile 1 stocks as firm size increases. A similar pattern obtains for other stock categories defined using B/M, price, and institutional ownership (results available upon request).

[27] The mean portfolio correlations for BSI measures across size quintile portfolios range from 0.541 to 0.666. The results are similar when we compute B/M-, price-, or institutional ownership-sorted quintile portfolios.

[28] When we use the market-wide BSI instead of portfolio BSI in the multi factor model, the results are qualitatively similar, but weaker, in terms of statistical significance. Specifically, for the quintile 1 size portfolio and other extreme portfolios considered in Section III (lowest price and institutional ownership, and highest B/M quintile portfolios), the market-wide BSI loadings are all statistically significant at the 10% level.

[29] For brevity, we do not report the detailed results from these correlation tests but they are available upon request.

[30] We also experiment with other regression specifications that contain lagged macroeconomic variables or innovations measured over the past k months (k = 1, 2, and 3). The results from these alternative specifications are very similar to the reported estimates.

[31] The correlation between raw and residual BSI provides another measure of the impact of innovations in macroeconomic variables on portfolio BSI. For size quintile portfolios 1 to 5, the correlations between raw and residual BSI time series are 0.903, 0.934, 0.903, 0.907, and 0.874, respectively. These high correlations indicate that innovations in macroeconomic variables are not primary drivers of portfolio BSIs.

[32] The correlation between market-wide BSI and ΔVWD is also positive (= 0.105).

[33] LST report that small stocks earn positive excess returns when closed-end fund discounts narrow. Our results show that the explanatory power of closed-end fund discounts for returns is not statistically significant in the 1991 to 1996 period. These results are consistent with Qiu and Welch (2004), who document a similar decline in the explanatory power of CEF discounts. A possible explanation, according to LST, is that institutional ownership has increased over time for all stocks, making the closed-end fund variable a noisier proxy for retail sentiment.

[34] Our results are consistent with Odean (1999), who finds a mean 252-day post-trade buy-sell return differential of −3.31% using an older version of the investor database.

[35] To obtain the risk-adjusted return differential, we construct a trading strategy. Each month, we take a long position in stocks that are purchased by our sample investors in month t, and we take a short position in stocks that are sold in month t. The portfolios are rebalanced at the beginning of each year. The four-factor alpha of this trading strategy provides a measure of the risk-adjusted performance differential.

[36] Consider an extreme case: in a given month, stock A has 999 buyers and 1 seller, while stock B has 1 buyer and no seller. According to our current BSI measure, the BSI of stock A is slightly lower than the BSI of stock B. Under the new BSI measure, the BSI will be the same (=1) for both stocks.

[37] To facilitate comparisons, we also summarize the BSI loading estimates for size-sorted portfolios.

[38] In DHS, idiosyncratic mispricing is completely arbitraged away while systematic mispricing remains due to limited arbitrage. Consequently, systematic factor sensitivity measures can be used as proxies for mispricing.

[39] For brevity, we only report the Granger causality probabilities, which indicate whether the lagged portfolio BSI can predict portfolio returns.

[40] For example, this variable serves as a proxy for information uncertainty in Zhang (2005) and Jiang, Lee, and Zhang (2005); similarly, in Baker and Wurgler (2005), it proxies for a stock's sensitivity to investor sentiment.

[41] The liquidity beta is estimated for each stock in the sample using the 1991 to 1996 returns data. The stocks with less than 24 months of data are excluded from the analysis.

[42] A curious difference between Kaniel, Saar, and Titman (2004) and our study is that they do not find a significant cross-sectional correlation in their measure of retail trade imbalance. This difference is probably due to difference in our research methodology (they use the principal component analysis while we use multiple, but related approaches, sampling frequency (they examine daily order imbalances while we examine monthly imbalances), aggregation mode (they use individual stock imbalances while we use portfolio level imbalances), data source (they use order flow data from the NYSE while our data contain both NYSE and NASDAQ stocks), and sample period (their data cover a more recent time period, 2000 to 2002, while our data span 1991 to 1996 period). In any event, to the extent that correlated retail trading affects stock returns, the phenomenon we document should be of broad interest even if it is not readily detectable in other data sources.

Table I

Summary Statistics: Retail Investor Trading Behavior

This table reports aggregate, stock-level, and investor-level trading statistics. Our sample consists of 62,387 retail investors that execute 1,854,776 trades in 10,877 stocks during the 1991 to 1996 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. We use 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 to obtain the proportions in Panels B and C respectively. The retail investor data are from a large U.S. discount brokerage house for the period 1991 to 1996.

Panel A: Aggregate Trading Statistics

Statistic	1991	1992	1993	1994	1995	1996
Number of buy trades	184,358	174,466	168,790	141,570	167,134	179,417
Number of sell trades	133,845	132,541	146,768	120,209	151,427	154,251
Average trade size (buys)	\$12,941	\$13,166	\$13,114	\$12,698	\$13,004	\$12,815
Average trade size (sells)	\$13,147	\$12,927	\$12,953	\$12,977	\$13,043	\$12,853
Total number of stocks traded	5,930	6,010	6,514	6,804	7,206	7,096
Total number of investors	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
At least 1 trade	56.94	54.16	57.85	55.98	58.63	60.29
5 or more trades	19.75	17.52	18.34	15.81	17.66	18.12
10 or more trades	10.36	9.21	8.77	7.26	8.53	8.98
25 or more trades	3.60	3.25	2.92	2.37	2.87	3.25
50 or more trades	1.45	1.32	1.22	0.89	1.07	1.38
75 or more trades	0.86	0.77	0.71	0.49	0.59	0.82
100 or more trades	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
At least 1 trade	21.91	20.00	23.17	27.06	36.61	44.72
5 or more trades	2.33	2.04	2.58	2.92	5.03	6.99
10 or more trades	0.56	0.47	0.60	0.68	1.37	2.17
25 or more trades	0.22	0.18	0.23	0.26	0.56	1.01
50 or more trades	0.11	0.09	0.11	0.12	0.29	0.55
75 or more trades	0.06	0.05	0.06	0.07	0.17	0.34
100 or more trades	0.01	0.01	0.01	0.02	0.03	0.07

Table II

Evidence of Common Directional Component in Retail Trading Activities

This table reports the simulation results (correlation statistics) from two sets of randomization tests that examine the existence of a systematic component in the trading activities of retail investors. In Panel A, we report the correlation statistics for randomly chosen stock portfolios. We form 1,000 pairs of non overlapping k-stock portfolios where $k = 50, 75, 100, 125, 250$, and 500. We then obtain the buy-sell imbalance (BSI) time series for these portfolios where the month-t BSI for portfolio p is defined as $BSI_{pt} = \frac{100}{N_{pt}} \sum_{i=1}^{N_{pt}} BSI_{it}$ and the month-t BSI for stock i 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. In Panel B, we report the correlation statistics for randomly chosen sets of investors. We form 1,000 pairs of mutually exclusive k-investor sets, where k = 500, 1,000, 1,500, 2,000, 2,500, and 5,000. The buy-sell imbalance (BSI) time series is obtained for these investor groups, where the month-t BSI for group i 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 retail investor data are from a large U.S. discount brokerage house for the period 1991 to 1996.

Panel A: Random Stock Portfolios

	Number of stocks in the random portfolio					
Statistic	50	75	100	125	250	500
Mean	0.234	0.367	0.461	0.534	0.600	0.612
Median	0.237	0.372	0.464	0.543	0.605	0.618
Std. Dev.	0.127	0.104	0.093	0.080	0.069	0.065
25^{th} *Pctl.*	0.142	0.296	0.399	0.484	0.556	0.566
75^{th} *Pctl.*	0.317	0.441	0.524	0.588	0.648	0.659

Panel B: Random Subsets of Investors

	Number of randomly chosen investors					
Statistic	500	1,000	1,500	2,000	2,500	5,000
Mean	0.147	0.210	0.272	0.326	0.365	0.496
Median	0.141	0.209	0.278	0.339	0.367	0.506
Std. Dev.	0.127	0.112	0.124	0.117	0.110	0.108
25^{th} *Pctl.*	0.068	0.133	0.188	0.252	0.295	0.436
75^{th} *Pctl.*	0.240	0.295	0.352	0.398	0.434	0.582

Table III
Category-Level Buy-Sell Imbalance (BSI): Time-Series Statistics and Correlations

This table reports the basic statistics and correlations (with standard risk factors) of the portfolio BSI time series for quintile portfolios obtained by sorting on size. The quintile portfolios are formed at the end of each year using the size break-points from the end of December. The portfolios are held constant throughout the following year. Panel A reports the basic statistics and Panel B presents the correlations. The retail investor data are from a large U.S. discount brokerage house for the period 1991 to 1996.

Panel A: Time-Series Statistics

Portfolio	Mean	Median	Std. Dev.	Min	Max
Size Quintile 1	3.48	3.25	9.32	−24.39	22.10
Size Quintile 2	0.42	−0.17	7.06	−14.08	15.74
Size Quintile 3	−3.29	−3.40	7.02	−18.74	12.01
Size Quintile 4	−6.68	−6.61	6.98	−22.64	8.87
Size Quintile 5	−10.60	−11.58	8.14	−25.82	9.18

Panel B: Correlations

Portfolio	RMRF	SMB	HML	UMD
Size Quintile 1	-0.141^{*}	0.215^{**}	0.196^{**}	-0.449^{***}
Size Quintile 2	-0.239^{**}	-0.148^{*}	0.135^{*}	-0.373^{***}
Size Quintile 3	-0.201^{**}	-0.192^{**}	0.036	-0.222^{**}
Size Quintile 4	-0.322^{***}	-0.286^{**}	0.144^{*}	-0.174^{*}
Size Quintile 5	-0.467^{***}	-0.127^{*}	0.130^{*}	-0.163^{*}

Table IV

Time-Series Factor Model Estimates for Size Portfolios

This table reports the factor model estimates for the five 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. We estimate the following time-series factor model:

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

Here, R_{pt} is the rate of return on the size-ownership portfolio, R_{ft} is the risk-free rate of return, $RMRF_t$ is the market return in excess of the risk-free 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 $t-2$, BSI_{pt} is the buy-sell imbalance for the size-ownership portfolio in month t, and ε_{pt} is the residual return on the portfolio. The portfolio BSI in month t is defined as $BSI_{pt} = \frac{100}{N_p}\sum_{i=1}^{N_p} BSI_{it}$, where BSI_{it} is the buy-sell imbalance of stock i in month t and 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. The retail investor data are from a large U.S. discount brokerage house for the period 1991 to 1996. The Newey-West adjusted t-values of the coefficient estimates are reported in parentheses.

Portfolio	Alpha	RMRF	SMB	HML	UMD	Portf. BSI	Adj. R^2
	0.463	0.867	1.448	0.694	−0.244		
Size	(1.825)	(10.188)	(9.895)	(5.634)	(−2.835)		0.814
Quintile 1	0.165	0.870	1.409	0.634	−0.134	**0.069**	
	(0.649)	(10.841)	(9.103)	(4.827)	(−1.540)	**(3.030)**	**0.857**
	−0.050	0.994	0.924	0.188	−0.166		
Size	(−0.490)	(27.356)	(21.787)	(4.573)	(−4.208)		0.958
Quintile 2	−0.058	0.995	0.926	0.188	−0.161	0.004	
	(−0.573)	(27.002)	(21.703)	(4.604)	(−4.021)	(0.335)	0.957
	−0.148	1.018	0.732	0.131	−0.018		
Size	(−2.063)	(46.190)	(24.334)	(4.097)	(−0.591)		0.974
Quintile 3	−0.140	1.020	0.734	0.132	−0.015	0.004	
	(−1.935)	(45.150)	(23.985)	(4.056)	(−0.536)	(0.508)	0.976
	−0.026	1.035	0.421	0.103	−0.039		
Size	(−0.367)	(46.599)	(14.454)	(4.515)	(−1.478)		0.959
Quintile 3	−0.018	1.036	0.422	0.103	−0.038	0.001	
	(−0.209)	(43.858)	(14.073)	(4.467)	(−1.458)	(0.148)	0.959
	0.011	1.053	0.009	0.067	−0.061		
Size	(0.246)	(50.614)	(0.510)	(3.371)	(−2.343)		0.982
Quintile 5	0.031	1.056	0.010	0.067	−0.060	0.002	
	(0.512)	(47.394)	(0.546)	(3.426)	(−2.303)	(0.426)	0.980

Table V
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 excess percentage 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 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, stock price, and institutional ownership variables. The retail investor data are from a large U.S. discount brokerage house for the period 1991 to 1996.

Panel A: Mean Percentage Excess Retail Ownership

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

Panel B: Mean Retail Trading Activity (NTA)

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

Table VI
BSI Loadings for Portfolios Formed on Firm Characteristics

This table reports the BSI loadings for portfolios formed on firm size, book-to-market, 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 constant throughout the following year. Panel A reports the portfolio break-points while Panel B reports the BSI loadings. We estimate the following time-series factor model:

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

Here, R_{pt} is the rate of return on the size-ownership portfolio, R_{ft} is the risk-free rate of return, $RMRF_t$ is the market return in excess of the risk-free 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 $t-2$, BSI_{pt} is the buy-sell imbalance for the size-ownership portfolio in month t, and ε_{pt} is the residual return on the portfolio. The portfolio BSI in month t is defined as $BSI_{pt} = \frac{100}{N_p} \sum_{i=1}^{N_p} BSI_{it}$, where BSI_{it} is the buy-sell imbalance of stock i in month t and 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. Panel C reports the Granger causality probabilities from a sentiment-return bivariate vector auto-regressive (VAR) model. An element in Panel C represents the impact of the column variable on the row variable. The retail investor data are from a large U.S. discount brokerage house for the period 1991 to 1996. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table VI(Continued)

BSI Loadings for Portfolios formed on Firm Characteristics

Panel A: Portfolio Break-Points

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

Panel B: Portfolio BSI Loadings

	Quintile Portfolios				
Sorting Variable	Low	Q2	Q3	Q4	High
Firm Size	0.069***	0.004	0.004	0.001	0.002
Book-To-Market	0.019	−0.015	0.009	0.020**	0.060***
Insti. Own.	0.071***	0.006	0.012	0.009	−0.030**
Stock Price	0.069***	0.055**	0.008	0.036	0.008

Panel C: Granger Causality Probabilities

	Quintile Portfolios				
Sorting Variable	Low	Q2	Q3	Q4	High
Firm Size	0.013	0.369	0.730	0.861	0.212
Book-To-Market	0.148	0.386	0.928	0.122	0.015
Insti. Own.	0.002	0.208	0.450	0.619	0.750
Stock Price	0.036	0.026	0.754	0.962	0.327

Table VII

Arbitrage Cost and Retail Sentiment Relation in Habitat Portfolios

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, that is, 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, and price. Panel B reports the portfolio BSI loadings for quintile portfolios sorted by arbitrage cost. The average number of stocks in a portfolio is reported in parentheses. To obtain the sentiment factor loading for a portfolio, we estimate the following time-series factor model:

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

Here, R_{pt} is the rate of return on the size-ownership portfolio, R_{ft} is the risk-free rate of return, $RMRF_t$ is the market return in excess of the risk-free 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 $t-2$, BSI_{pt} is the buy-sell imbalance for the size-ownership portfolio in month t, and ε_{pt} is the residual return on the portfolio. The portfolio BSI in month t is defined as $BSI_{pt} = \frac{100}{N_p} \sum_{i=1}^{N_p} BSI_{it}$, where BSI_{it} is the buy-sell imbalance of stock i in month t and 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. The retail investor data are from a large U.S. discount brokerage house for the period 1991 to 1996. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Mean Arbitrage Cost Estimates

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

Panel B: Portfolio BSI Loadings

	Arbitrage Cost Quintile Portfolios				
Habitat	Low	Q2	Q3	Q4	High
Small Firm Size	0.001(350)	0.014(519)	0.018(723)	0.047**(850)	0.081***(804)
High Book-To-Market	−0.025(179)	−0.007(298)	0.006(339)	0.074***(267)	0.069**(168)
Low Insti. Own.	0.006(243)	0.005(84)	0.020*(96)	0.041**(145)	0.077***(228)
Low Stock Price	−0.005(7)	0.011(43)	0.005(167)	0.095***(426)	0.093***(602)

Panel C: Portfolio BSI Loadings with Liquidity Beta as Arbitrage Cost Proxy

	Liquidity Beta Quintile Portfolios				
Habitat	Low	Q2	Q3	Q4	High
Small Firm Size	0.106***(583)	0.066***(510)	0.061***(439)	0.048(463)	0.008(611)
High Book-To-Market	0.054***(321)	0.010(280)	0.013(227)	−0.026(236)	−0.012**(341)
Low Insti. Own.	0.038***(195)	0.022***(145)	0.013**(248)	0.014(200)	0.006(178)
Low Stock Price	0.132***(426)	0.054**(163)	0.068**(129)	0.029(153)	0.017(412)

Table VIII
BSI Loading Estimates for Double-Sorted Portfolios: Firm Characteristics and Arbitrage Costs

This table reports the BSI loading estimates for portfolios formed on firm size, book-to-market (B/M), level of institutional ownership (IO), month-end stock price, and arbitrage cost. 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. Quintile portfolios are formed at the end of each year using break-points from the end of December. Panels A-D report the portfolio BSI loadings for quintile portfolios sorted on arbitrage cost and size, B/M, IO, and stock price, respectively. The average number of stocks in a portfolio is reported in parentheses. To obtain the sentiment factor loading for a portfolio, we estimate the following time-series factor model:

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

Here, R_{pt} is the rate of return on the size-ownership portfolio, R_{ft} is the risk-free rate of return, $RMRF_t$ is the market return in excess of the risk-free 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 $t-2$, BSI_{pt} is the buy-sell imbalance for the size-ownership portfolio in month t, and ε_{pt} is the residual return on the portfolio. The portfolio BSI in month t is defined as $BSI_{pt} = \frac{100}{N_p} \sum_{i=1}^{N_p} BSI_{it}$, where BSI_{it} is the buy-sell imbalance of stock i in month t and 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. The retail investor data are from a large U.S. discount brokerage house for the period 1991 to 1996. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Arbitrage Cost and Size Sort

	Size Quintile				
Arb Cost Quintile	Small	Q2	Q3	Q4	Large
Low	0.001(350)	0.011(242)	0.014(208)	0.017(236)	0.011(320)
Q2	0.014(519)	0.007(252)	−0.001(211)	−0.020(190)	−0.019(121)
Q3	0.018(723)	0.010(232)	−0.025(142)	−0.014(95)	−0.031(37)
Q4	0.047**(850)	0.024*(148)	−0.056**(74)	−0.050*(33)	−0.055(19)
High	0.081***(804)	0.040**(63)	−0.059**(21)	−0.050(15)	−0.040(11)

Panel B: Arbitrage Cost and B/M Sort

	B/M Quintile				
Arb Cost Quintile	Growth	Q2	Q3	Q4	Value
Low	−0.000(140)	−0.004(158)	−0.004(194)	0.002(229)	−0.025(179)
Q2	−0.029(165)	−0.009(178)	−0.010(183)	−0.013(194)	−0.007(298)
Q3	−0.022(182)	−0.025(166)	0.005(146)	−0.001(164)	0.006(339)
Q4	0.032(251)	−0.016(153)	−0.017(120)	0.014*(124)	0.074***(267)
High	0.048**(297)	0.029(107)	0.058***(79)	0.029**(74)	0.069***(168)

Table VIII(Continued)

BSI Loading Estimates for Double-Sorted Portfolios: Firm Characteristics and Arbitrage Costs

Panel C: Arbitrage Cost and Institutional Ownership Sort

	IO Quintile				
Arb Cost Quintile	Low	Q2	Q3	Q4	High
Low	0.006(243)	0.019(145)	0.000(163)	0.002(221)	−0.015*(283)
Q2	0.005(84)	−0.024(152)	−0.003(205)	0.007(267)	−0.028***(332)
Q3	0.020*(96)	0.009(183)	−0.004(226)	0.026*(231)	−0.026(214)
Q4	0.041**(145)	0.010*(231)	0.014**(185)	0.027(136)	0.117***(80)
High	0.077***(228)	0.039**(152)	0.020*(60)	0.068**(35)	0.036**(12)

Panel D: Arbitrage Cost and Price Sort

	Price Quintile				
Arb Cost Quintile	Low	Q2	Q3	Q4	High
Low	−0.005(7)	−0.007(80)	0.006(239)	0.011(349)	0.004(685)
Q2	0.011(43)	−0.004(170)	−0.012**(281)	0.006(400)	−0.025**(399)
Q3	0.005(167)	0.050(346)	0.020*(305)	−0.005(242)	−0.021(174)
Q4	0.095***(426)	0.110**(339)	0.032*(184)	−0.029(111)	−0.036*(60)
High	0.093***(602)	0.100***(196)	0.017**(71)	0.016(27)	−0.020(9)

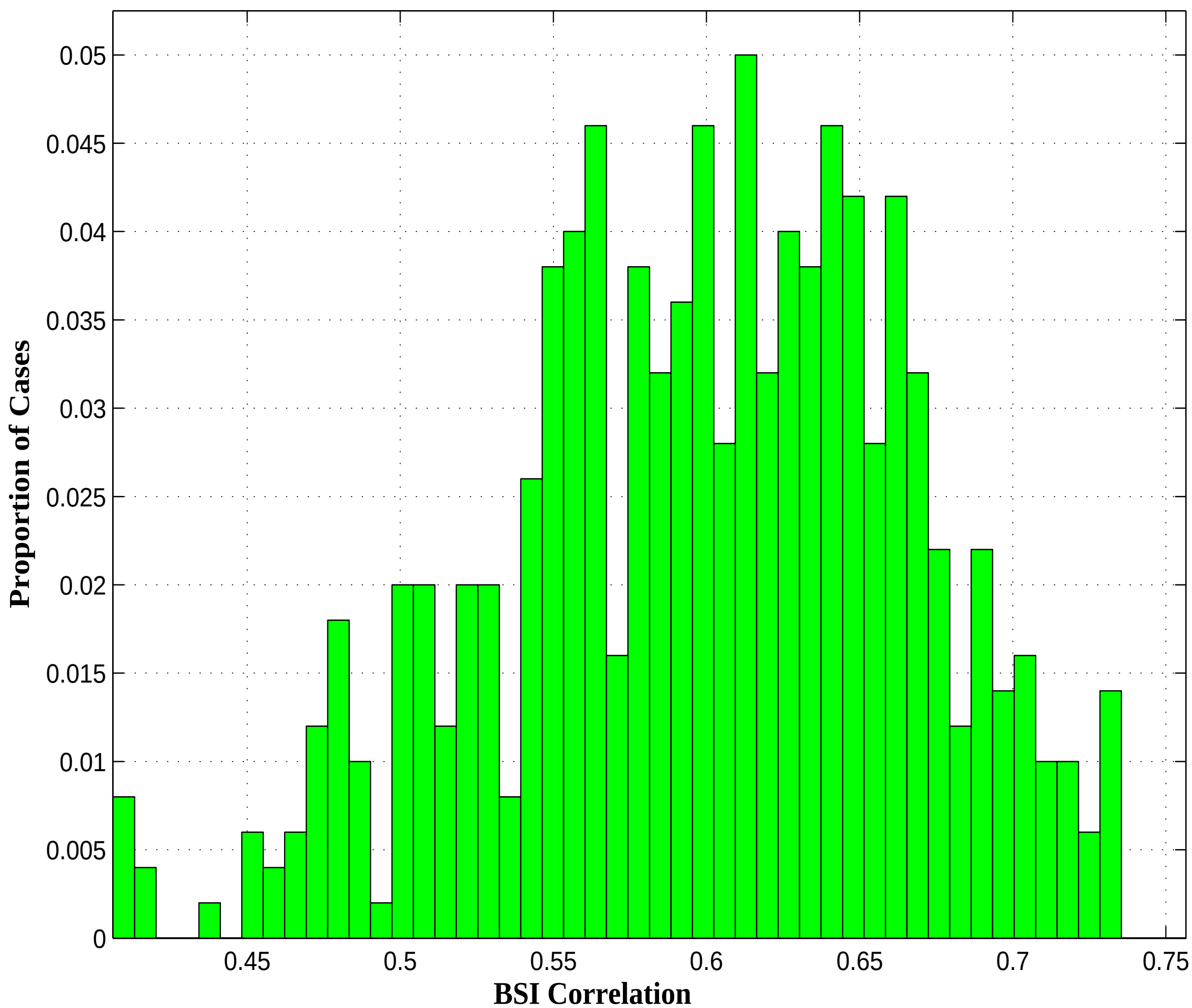

Figure 1. **Buy-sell imbalance (BSI) correlation distribution.** This figure shows the residual BSI correlation distribution obtained by forming 1,000 pairs of non overlapping 250-stock random-portfolios. The BSI of portfolio p in month t is defined as $BSI_{pt} = \frac{100}{N_p} \sum_{i=1}^{N_p} BSI_{it}$, where the BSI for stock i in month t is defined as $BSI_{it} = [\sum_{j=1}^{D_t}(VB_{ijt} - VS_{ijt})]/[\sum_{j=1}^{D_t}(VB_{ijt} + VS_{ijt})]$. 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. The retail investor data are from a large U.S. discount brokerage house for the period 1991 to 1996.

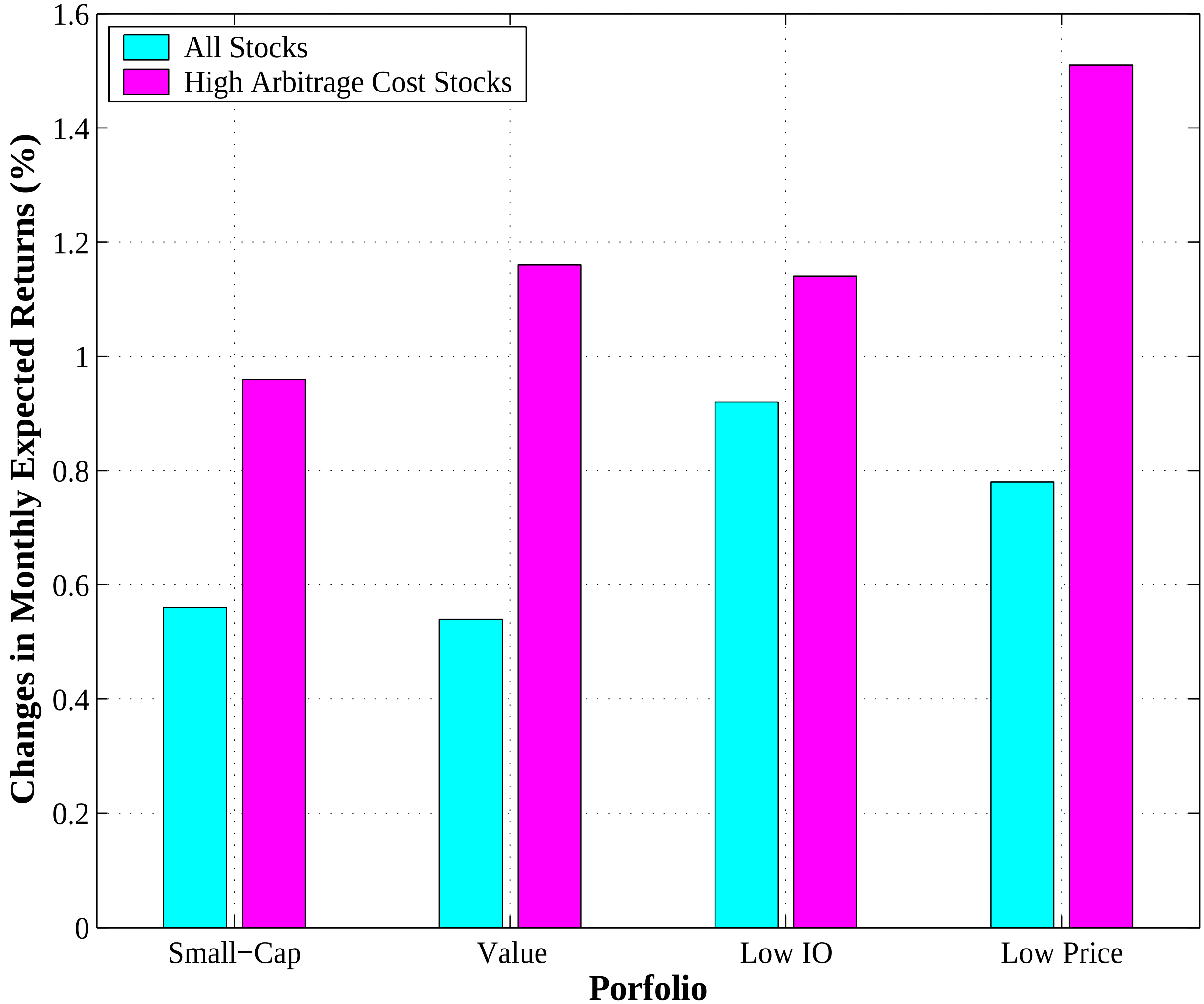

Figure 2. **The economic significance of changes in BSI for various stock portfolios.** This figure depicts the change in monthly expected returns associated with a one-standard deviation change in the residual BSI for various stock portfolios. The residual BSI is the residual from a regression of the portfolio BSI on a vector of standard risk factors – $RMRF$, SMB, HML, UMD, and $MACRO$, where $MACRO$ is a vector of innovations in macroeconomic variables (i.e., UI, MP, ΔTS, ΔRP, $\Delta UNEMP$, and $\Delta WAGE$). We report results for stocks in the quintile with the highest retail ownership (e.g., Small-Cap, Value, Low Institutional Ownership (IO), and Low Price). The left bar represents the change in expected monthly returns for all firms in a given quintile; the right bar represents the result for firms that are also in the highest arbitrage cost quintile. We use the variance of the residual from a capital asset pricing model (CAPM) regression, that is, the idiosyncratic risk of each firm, as a proxy for arbitrage costs (Wurgler and Zhurvaskaya (2002)). The retail investor data are from a large U.S. discount brokerage house for the period 1991 to 1996.