Does Sign Matter More than Size?
An Investigation into the Source of Investor Overconfidence

Sankar De[a], Naveen. R. Gondhi[b] and Bhimasankaram Pochiraju[c]

Present version: November, 2010.

Abstract

Using a unique and large dataset, the present paper contributes new insights to the growing literature on behavioral biases in portfolio decisions of individual investors. We find that the sign of the outcomes of recent past stock trades, where a positive sign indicates a profitable trade and a negative sign an unprofitable trade, influences the current trading decisions of investors strongly. Further, the influence of the sign of past trades is significantly stronger than the size of the gains or losses from the same trades. We also find that, on an average, trading under the influence of the sign of past trades consistently results in decline in profits for the investors. Our findings are consistent with a behavioral explanation suggested by recent research in experimental psychology that the investors are more sensitive to the affective intensity of a stimulus than to its magnitude. The effects of the behavioral pattern in question are economically large as well as statistically significant. We introduce several methodological innovations in this paper.

Key words: individual investors, institutional investors, trading behavior, overconfidence, rationality.

[a] Centre of Analytical Finance (CAF), Indian School of Business, Hyderabad 500032, India.
[b] Centre for Analytical Finance (CAF), Indian School of Business, Hyderabad 500032, India.
[c] Corresponding author. Centre of Analytical Finance (CAF), Indian School of Business, Hyderabad 500032, India. Email: Bhimasankaram_Pochir@isb.edu

The authors thank National Stock Exchange for the use of their proprietary trading data, and Citi Foundation for financial support. They also thank Sumit Agarwal, Michael Brennan, Sanjiv Das, Ankit Jain, NR Prabhala, Subrata Sarkar, Manpreet Singh, and seminar participants at Indian School of Business, National Stock Exchange, and Tel Aviv University for helpful suggestions. The authors remain responsible for all errors in the paper.

Does Sign Matter More than Size? An Investigation into Source of Investor Overconfidence

Present version: November, 2010

Abstract

Using a unique and large dataset, the present paper contributes new insights to the growing literature on behavioral biases in portfolio decisions of individual investors. We find that the sign of the outcomes of recent past stock trades, where a positive sign indicates a profitable trade and a negative sign an unprofitable trade, influences the current trading decisions of investors strongly. Further, the influence of the sign of past trades is significantly stronger than the size of the gains or losses from the same trades. We also find that, on an average, trading under the influence of the sign of past trades consistently results in decline in profits for the investors. Our findings are consistent with a behavioral explanation suggested by recent research in experimental psychology that the investors are more sensitive to the affective intensity of a stimulus than to its magnitude. The effects of the behavioral pattern in question are economically large as well as statistically significant. We introduce several methodological innovations in this paper.

Key words: individual investors, trading behavior, sign and size, affect and scope, rationality.

JEL classification: D19, G14

I Introduction

Recently, a growing field of research in behavioral finance has addressed the task of documenting and understanding how cognitive biases of various kinds influence portfolio decisions of individual investors by observing their actual market behavior. Using a large and unique dataset, the present paper contributes new insights to this literature. First, the sign of the outcomes of of recent past stock trades, where a positive sign indicates a profitable trade and a negative sign an unprofitable trade, influences the current trading decisions of investors strongly. Further, the influence of the sign of past trades is significantly stronger than that of the size of the gains or losses from the same trades. For example, following a trade yielding a $500 profit and another a $5,000 profit, the next trade of an individual investor is influenced more by the fact that the outcome in each case was profitable than by the fact that the outcome was considerably larger in the second case. A similar result is observed for unprofitable trades with different magnitudes of loss. We also find that, on an average, trading under the influence of the sign of past trades consistently reduces profits for the investors and that, further, stronger the influence of sign, lower are the profits from current trades. The last result indicates that the observed trading pattern is not driven by "rational" reasons, but is behaviorally motivated. Interestingly, the effect of the size of past outcomes on current profits is positive. However, the negative effect of sign overwhelms the positive effect of size, resulting in a negative association between past profits and current profits overall.

As we have indicated above, our finding that trading under the influence of the sign of past outcomes reduces current profits consistently suggests a non - rational explanation for the observed behavior pattern of the investors. Indeed, a stream of recent research in experimental psychology provides a suitable explanation. This research examines the relationship between the magnitude of a stimulus and its value to the subjects receiving the stimulus, and considers two distinct psychological processes to arrive at the value (see Hsee and Rottenstreich, 2004; Hsee, Rottenstreich, and Xiao, 2005). One of them is deliberate, rule - based, and can be even algorithmic. In its deliberateness, it can approach rational decision - making behavior in traditional economic models. The other process is associative and affect - based. It is an essentially subjective process. Hsee and Rottenstreich (2004) characterize the two processes as valuation by calculation and valuation by feeling. The two processes are not necessarily mutually exclusive. In a given case, the two can be combined to yield a concave value function of the kind common in economic analysis where constant increments of the size of an outcome yield successively smaller increments of value of the outcome for the subjects. This theory accounts for an amply documented finding in experimental psychology that people become less sensitive to the magnitude of a stimulus as the magnitude increases. However, when feeling predominates, it can induce extreme concavity of the value function,implying extreme size-insensitivity. In this situation, value is highly sensitive to the presence or absence of a stimulus (that is, a change from zero to some scope) but is largely insensitive to further variations in scope. Our analysis, based on actual trading data rather than experimental evidence, leads us to a similar finding. An increase in profits that moves an investor from the domain of losses to the domain of gains has a stronger influence on her trading activity than equivalent increases in profits that are entirely in the domain of gains or losses[1]. We also consider several alternative rational explanations for our findings, including trading by investors who are learning about their true investing skills, private information, portfolio rebalancing, and wealth effects of past profits. None of them, we find, explains our results (see section VIII below).

Our findings not only provide new insights into investor behavior, they also throw new light on behavioral patterns documented in the existing literature, such as excessive trading by overconfident individual investors (Barber and Odean 1999,2000, 2001). The findings suggest that the sign of the outcomes of past trades, rather than their dollar size, induces overconfidence in individual traders. Existing studies on investor

[1]We thank Terry Odean for this particular articulation of our intuition.

overconfidence have documented or predicted a positive association between past outcomes and current trading activities of investors subject to overconfidence (Barber and Odean 2001; Gervais and Odean, 2001; Linnainmaa, 2009). However, what causes investor overconfidence remains an open and largely unanswered question. Certain hypotheses including characteristics of individual investors such as gender (Barber and Odean, 2001) and psychological traits such as an attribution bias leading to assuming responsibility for successes but not for failures (Odean 1999; Barber and Odean 1999; Gervais and Odean 2001), have been suggested in the existing literature. It is difficult to provide direct tests of the latter hypothesis. In our study we provide an alternative hypothesis behind investor overconfidence which is amenable to direct testing. Our tests consistently detect a positive relation between past profits and current investment decisions of individual investors and establish that this positive association is driven more by the success or failure of the past trades, that is their sign, than by their magnitude. The observed investor behavior consistently results in losses and, therefore, cannot be rational. We probe further and find that, in fact, size and other aspects of past profits have a *positive* effect on current profits. However, the sign of past outcomes has a pronounced negative impact on current profits which overwhelms the positive impact of size and results in an overall negative association between past profits and current profits.

Our results hold after allowing for an array of controls, including observed investor-specific characteristics such as past trading records (including number of trades and trading performance), at the same time using individual fixed effects to control for all unobserved characteristics and correcting for contemporaneous correlations between trading activities of individual investors. We also control for market movements and day-to-day variations in market-wide trading activity. Further, we conduct each test with several alternative specifications (different time-lags over which the effect of a past outcome remains potent; different holding periods for estimating gains/losses from trades; different sample sizes) as well as with alternative methodologies (GLS regressions; probabilistic regressions). The main results remain unchanged.

Our results are not only statistically robust, but also economically significant. The effects of the sign of the past outcomes on current trading decisions are economically large. The current trading volume is 7.7% more for the successful traders than for the losers in recent past trades, after controlling for the size of the respective gains and losses. Likewise, the probability that a successful trader will increase her current trading volume is 1.7% more than for a loser. On the other hand, profits from current trades is almost 60% less for the recent winners than for the losers. The differences are statistically significant in each case at 1% level. As a final verification, we confirm our findings by using the procedure suggested by Kruskal (1987) to estimate the relative importance of different explanatory variables in a regression model (see section VI below). While past profits or losses on an average explain about 48% of the variation in the current trading volume of the individual investors after controlling for all other relevant variables, the sign of the past outcomes accounts for about three-fourths of this variation while size accounts for less than one-fourth.

In the process of arriving at our results, we introduce several methodological innovations that, we believe, add strength to our analysis. Some of the innovations are potentially useful for other empirical studies and represent new contributions by themselves. It is not straightforward to separate the respective shares of the sign and the size of a past outcome in the total impact that the outcome exerts on the current trading decisions of an individual investor. If we include the sign of the outcome and also the outcome itself, whether a profit or a loss, simultaneously as explanatory variables in a regression model, we end up underestimating the true effect of the sign. The estimated regression coefficient will reflect the residual effect of the sign, if any, over and above the effect of the outcome. On the other hand, if we include only the sign and not the actual outcome of the past trade, then the effect of the sign is likely to be overestimated. Since the sign is likely to be highly correlated with the outcome of the past trade (the correlation coefficient is 0.943 for our sample), the regression coefficient for the sign variable may very well capture a part of the impact of the outcome in addition to its own effect. Hence, we are required to decompose the outcome of a past trade into its sign and another component which will reliably capture the residual effect of the past outcome on the investor's current trading decisions. To accomplish this task, we use an innovative three-step procedure (see section III below). For another innovation, we consider variable holding periods for identifying gains/losses for the investors in our sample. Until now, a fixed holding period has been used in the literature on investor overconfidence[2]. (see, for example, Odean, 1999; Barber and Odean, 1999, 2000, 2001; and Statman *et al*, 2006). However, a fixed holding period requires the researcher to ignore all trades in the same security (e.g. from buying to selling the same security) completed within the holding period. Under our method, to assess the profitability of a trade we compare the transaction price to the price at the time a particular investor's next trade. Apart from not losing any transactions, this method is built on the rationale that the investor uses the latest information to evaluate her performance in the last trade. This shapes the psychological reinforcements that she takes to her next trade. Clearly, this method of computing trading performance involves different holding periods for different transactions.

The sample used in the present study is very large. Our sample includes nearly 106 million transactions with a monetary value of almost 1825 billion rupees made by about 0.99 million individual investors. Our dataset includes all transactions during January - June 2006 in the fifty stocks that constitute the S&P CNX Nifty index. S&P CNX Nifty is a well-diversified stock index covering a broad spectrum of the Indian economy. Nifty stocks represented about 58% of the total market capitalization of all stocks registered with the National Stock Exchange of India (NSE) on the last trading day of 2006 (Dec 29, 2006). NSE generated the third largest number of trades among all exchanges in the world, and the largest number among all electronic order - book markets in 2006 and in 2008, the last year for which this information is available. By

[2]Grinblatt and Keloharju, 2001, also use a variable holding period for buys; they do not consider sales in their study

volume, its position was 18^{th} among all exchanges in both years. Section II of this paper describes the data in detail. We note here that, combining trading data publicly available from the NSE (such as execution date and time, trade size, price etc.) with non-public data that the NSE made available to the authors of the present study, we are able to track all transactions made by each investor in our sample.

The rest of the paper is organized as follows. Section II describes our data and the empirical variables used in our tests. In section III, we discuss the methodological innovations that we introduce to conduct the tests. Section IV presents the results of the tests with trading volume of individual investors as their decision variable. In section V, we investigate the effects of the observed behavioral patterns in trading on investor wealth. In section VI, we conduct additional tests to confirm the important role of the sign of past outcomes in current trading decisions. In section VII, we examine several alternative explanations for our findings, including trading driven by investors learning about their true investing skills, private information, portfolio rebalancing, and wealth effects of past profits. In section VIII, we present our conclusions.

II Data and variables

II.1 *Data source*

There are two major stock exchanges in India: Bombay Stock Exchange (BSE), and the NSE. Established in 1875 as a stockbrokers' association, the BSE is the oldest stock exchange in Asia. Today, however, the majority of stock trading and almost all trading in stock derivatives take place on the NSE. Established in 1994, the exchange is a limited liability company owned by public sector financial institutions. Brokerage firms do not own the exchange, and are represented neither on the board of directors nor in the management team. They are franchisees of the exchange, and are represented through exchange-appointed committees.

Stock trading on NSE commenced in November 1994. NSE is organized as an order-driven and open electronic limit order book market. NEAT, an automated screen-based trading system, enables members from across India to trade anonymously with one another on a real-time basis. NEAT operates on strict price time priority. The types of orders and books that exist internationally in order-driven markets typically exist also on NSE, including limit orders, market orders, hidden orders, disclosed quantity orders, regular and odd-lot orders, stop-loss orders, upstairs market, negotiated trade book etc. NSE generated the third largest number of trades among all exchanges in the world, and the largest number among all electronic order-book markets in 2006 (the year of our sample). In 2009, the last year for which this information is available, its position was fourth among all world exchanges. By volume, its position was 18^{th} in 2006 as well as in 2009.

II.2 *Data description*

The trading data used in the present study pertain to the 50 stocks included in S&P CNX Nifty index during January - June, 2006. The sample period included 123 trading days. S&P CNX Nifty is a well-diversified fifty-stock index covering a broad spectrum of the Indian economy. Nifty stocks represented about 58% of the total market capitalization of all NSE stocks on the last trading day of 2006 (Dec 29, 2006).

NSE maintains exceptionally high-quality trading records. The data used in this study come from two separate NSE datasets. The intra-day trades data for each exchange-registered stock for variables such as execution date and time, price, trade volume, etc. are publicly available from the NSE at a moderate cost. The second dataset is not publicly available. It contains information about orders, including submission date and time, limit price, size, order type as well as client identification codes and category information. The information from the two datasets was sufficient for us to identify all trades made by a given investor, and all orders placed by the said investor, on each trading day over the entire sample period in each stock in our sample.

The data provided by NSE classifies all investors into 17 categories. For our analysis, we focus on the trading activities of the individual investors. Note from table 1 below that 0.99 million individual investors traded during the sample period. They accounted for 98.5% of all investors. They were also the biggest category of investors with respect to the volume of trading, accounting for 1825.67 billion, as well as the number of transactions, placing 106 million trades during the sample period.

(Table 1 here)

Table 1 also indicates the gains and losses of the individual investors during the sample period. The gains and losses were computed using a variable holding period after correcting for stock dividends and stock splits. Note that, as a group, the individual investors lost a total of 7.66 billion to other investor categories (mostly to the institutional investors). The table indicates the proportion of winning/losing transactions as well as winning/losing trading days for the individual investors. As we explain below, a trading day is the basic unit of time in our analysis.

Since the objective of the present study is to investigate if and how the outcomes of past trades impact current trading behavior of individual investors, we can only consider investors with a minimum of two observations. The final samples used in this study includes investors who traded on at least two days out of the 123 trading days in the sample period. From table 1, during our sample period 80.9% of the individual investors traded on at least two days. For robustness checks, we run our tests with different inclusion cut-offs (at least six and eleven observations, in other words trading records on at least five and ten trading days prior to the current day). The results were qualitatively the same.

II.3 *Variables*

Table 2 below describes the variables we use in our regression models.

(Table 2 here)

The regression models we estimate in this study include the volume (dollar value) of trades placed by an investor in the current period as the dependent variable. Volume of trading by investors, given by $Volume_{it_r}$ for investor i on trading day t_r, aggregates all trades, including buys and sales, on the day. The variable is a more appropriate indicator of an investor's trading activity than net investments which nets out sells from buys (or vice versa), since the independent variables in our regression models include profits from both buys and sales. Our approach is consistent with Barber and Odean (1999, 2000, 2001)who use the sum (average)of purchase and sales turnovers as the dependent variable in their tests.

The independent variables of interest in the regression models are either recent past profits of the investors in our sample, indicated by $RProfit_{it_r}(k)$ for investor i on trading day t_r, or the $Sign$ and $Residual$ variables for the same profits. In the next section of this paper, we discuss the methodology we employ to separate the $Sign$ and the $Residual$ of an outcome. k indicates the number of lags used to compute recent past profits, and is usually one (except when we run robustness tests with k = 5). In other words, recent past profits for an investor comprise profits on the investor's last trading day.

The control variables in our regression models include investor-specific controls, market controls, and controls for time periods. Among investor-specific controls, we use cumulative profits from the investor's distant past trades given by $DProfit_{it_r}(k)$. By construction, the variable excludes profits over the recent k trading days. When k = 1, distant past profits include profits on all but the last trading day. A profit on past trades makes more funds available for trading today, and avoids the transaction costs of raising the funds from external sources. Besides, the wealth effects of successful past trades may increase appetite for risk and trading (Kaustia and Knupfer, 2008). These considerations make $DProfit_{it_r}(k)$ an appropriate control variable for investor i.

Profit for investor i from a transaction included in either $RProfit_{it_r}$ or $DProfit_{it_r}$ reflects how much better (worse) off the investor is as a result of a past transaction when she is ready to make the next trade in the same stock. The rationale for our profit estimation method is that the investor is very likely to evaluate her performance in the last trade when she is ready to trade again. Her performance at that point in time should shape the psychological reinforcements that she takes to her next trade. While realistically estiamting the impact of past trades on current trading decisions, our proft estimation method introduces a couple of novel features. First, the estimated profits reflect the *actual* period between trades for an investor rather than an arbitrarily fixed holding period. The method lets the trading records of the investors in our sample indicate the holding periods which are often different for different transactions for a given investor. Second, we estimate profits whether the last transaction is a buy or a sale. Profit for investor i from a buy transaction is calculated as the difference between the price at the time of the next trade in the same stock (either a sale or a new buy) and the buy price. On the other hand,profit from a sale transaction is calculated as the difference between the sale price and the price at the time of the next trade in the same stock (either a buy, or a sale if the investor did not sell all stocks the previous time). Though the profit in the latter case is counterfactual, and is usually ignored in studies in investor behavior[3]. , in our view it is important to take it into account because it determines how much better or worse off the investor feels on the next trading day because she sold out on a previous day. If the price has gone down over the interval, she should feel vindicated and perhaps more overconfident. On the other hand, if the price has gone up, her behavioral reaction could be quite the opposite. In other words, it is a reasonable hypothesis that counterfactual profits will influence the investor's next trade. Of course, the hypothesis should be tested. In the case of the present study, inclusion of counterfactual profits in the total profits of an investor is warranted for another reason. If counterfactual profits really do not matter, then their inclusion will make the regression results of the impact of past profits on current trading decisions weaker to some extent. In other words, the reported results will be biassed against our

[3] A notable exception is Odean *et al* (2010).

hypothesis. On the other hand,if counterfactual profits indeed matter, they should of course be included.

Please note that in section VI of this paper below, in addition to a variable holding period described above, we use a 30-day holding period to examine the profitability of sign-influenced trading strategy. We do so in order to compare our results to those in the existing literature which have mostly used a similar fixed-holding period to assess profits.

Our final investor-specific control is $Activity_{it_r}$. In our framework the cumulative number of trades an investor has placed prior to her current trading day indicates her current activity level. The inclusion of this control variable in our test models is, of course, easily justified. A number of recent studies have found that, by and large, with more trading, behavioral biases in trading diminish. The $Activity_{it_r}$ variable should be correlated with,though perhaps not identical with, investor experience. We note that the choice of an appropriate proxy for experience is not a settled issue in the existing empirical literature on investor behavior. Several variables have been used, depending on data availability and other factors, including investor age (Choi *et al*, 2008), years of trading experience (Seru *et al*, 2009), cumulative number of trades placed by an investor (Seru *et al*, 2009) etc. As we have said above, the primary objective of our exercise in this paper is to investigate if and how the outcomes of past trades affect current trading behavior of investors. The behavioral effects that are germane to our exercise are likely to come only from personal experience of trading.[4].

The regression models also include market controls, such as the risk-free rate and market portfolio return to control for market movements, and day-of-the week dummies, $Daydummy_1 - Daydummy_4$, to control for the day of the week effect, with Monday as the base day. Additionally, depending on the particular test, we include other control variables.

[4]We note that, though Seru *et al* (2009) use both cumulative number of trades and years of experience as a proxy for experience, their findings indicate the superiority of the former measure.

Table 3 below presents the summary statistics of the relevant variables used in regression: $Volume$, $Rprofit(k)$ for $k = 1$ and 5, $DProfit(k)$ for $k = 1$ and 5, and $Activity$ for the investors in our sample.

(Table 3 here)

In view of of the skewed distribution, we winsorize our sample by 1% at the top and the bottom for all relevant variables: $Volume$, $RProfit$, $DProfit$, and $Activity$. However, to verify robustness of our results, we perform all our tests in this paper with full as well as winsorized samples.

III Methodological issues

III.1 *Sign and size of past profits*

We wish to examine the impact, if any, of the sign of the outcome of past trades on an investor's current trading decisions. However, any observable impact of past outcomes may also be due, in part if not in entirety, to other aspects of past profits or losses besides sign. Hence, we are required to decompose the outcome of a past trade into its sign and another component which will reliably capture the residual effect of the past outcome on the investor's current trading decisions. Henceforth, we label the second component "residual". We expect the residual variable to represent primarily the dollar size of the gain or loss from the past trade, size being the other important aspect of an outcome besides its sign. Therefore, we use the two terms, residual and size, interchangeably, if necessary.

It should be obvious at this stage that the appropriate methodology to determine the residual or size as defined above, in isolation from the sign, of the outcome of a trade may not be straightforward. However, decomposing the effect of outcomes of past trades into sign and residual effects is crucially important for our analysis. Including the sign and residual in the same regression to explain the variation in the current trading activities of the investors allows us to compare the two effects, and determine which effect is more important. We discuss below our methodology to separate the two effects.

For investor i on trading day t_r, we define $Sign_{it_r}(k)$ as a binary variable taking value 1 if $RProfit_{it_r}(k)$ is positive, and zero otherwise[5]. In other words, for the variable to take a value of one, the combined profits for the investor for the previous k trading days must be positive.

In order to determine the residual variable, $Residual_{it_r}(k)$ for investor i on trading day t_r, when the trading decision is trading volume, we use a three-step procedure consisting of the following three regression models:

$$\begin{aligned} Volume_{it_r} &= \alpha_i + \beta_1 RProfit_{it_r}(k) + \beta_2 Activity_{it_r} + \beta_3 DProfit_{it_r}(k) + \beta_4 Rf_{t_r} + \beta_5 Index_{t_r} + \\ & \sum_{g=1}^{4} \beta_{5+g} Daydummy_{t_r g} + \epsilon_{1it_r} \qquad (1) \\ Volume_{it_r} &= \gamma_i + \theta_1 Sign_{it_r}(k) + \theta_2 Activity_{it_r} + \theta_3 DProfit_{it_r}(k) + \theta_4 Rf_{t_r} + \theta_5 Index_{t_r} + \\ & \sum_{g=1}^{4} \theta_{5+g} Daydummy_{t_r g} + \epsilon_{2it_r} \qquad (2) \end{aligned}$$

$$e_2 = f(e_1) + \epsilon_3 \qquad (3)$$

where e_1 and e_2 are the residuals from models (1) and (2) respectively, and $f(e_1)$ is an appropriate regression function. The exact form of the function will be driven by data.

Model (1) estimates the impact of $RProfit$ on $Volume$, subject to the control variables (please see the previous section for the justification of the controls). The residual from the regression model (1), e_1, comprises the impact of all missing variables and noise. In the second step, model (2) estimates the impact of $Sign$ on $Volume$, subject to the same control variables. Therefore, the residual from model (2), e_2, comprises the impact of $RProfit$ sans $Sign$, all missing variables and noise. In the last step, us-

[5]Note that, instead of (1,0), if the binary variable $Sign$ takes any ordered pair (a,b) where a > b, the sign of the regression coefficient and the corresponding t-statistic will remain the same.

ing model (3) we regress the residual from model (2) on the residual from model (1). Hence, the residual from model (3) represents the impact of $RProfit$ sans $Sign$. The residual from model (3) is henceforth called $Residual$. Note that its unit is Indian currency (INR), since the error terms in regression models (1) and (2) are also in Indian currency.

Since the error term in model(2) includes the variable $Residual$, there could be an endogeneity issue in model (2) if the correlation between the $Sign$ and $residual$ variables were appreciable. We verify that the actual correlation for our sample is as low as 0.1. Further, since the error term in model (2) contains the missing variables included in the error term in model (1) plus $Residual$, the relationship between the residuals from the two regressions (1) and (2) is in one direction only. Hence, there is no endogeneity problem due to reverse causation in model (3).

Anscombe (1967) and Mosteller and Tukey (1977) have also carried out regression of the residual from one regression on the residual from another. However, their context as well as procedure was quite different. Their objective was to assess the effectiveness of bringing a new regressor, in addition to several existing regressors, into a regression equation (added variable plot). Further, in their case, the regression coefficient estimate in their residual vs residual regression is important, whereas we are interested in the residual from our residual vs residual regression.

III.2 *Alternative holding periods*

As we have discussed in the preceding section, our profit estimation method lets the trading records of the investors in our sample indicate the holding periods which are often different for different transactions for a given investor. The method captures how much better (worse) off a given investor is as a result of a transaction when she is ready to trade again in the same stock. The rationale for our profit estimation method is that the investor is very likely to evaluate her performance in the last trade at the time of her next trade in the same stock. The evaluation will shape the psychological reinforcements that she takes to her next trade. Under this method, the estimated profits reflect the *actual* period between trades for an investor rather than an arbitrarily fixed holding period. However, in section VI of this paper below, in addition to a variable holding period described above, we use a 30-day holding period to examine the profitability of sign-influenced trading strategy. We do so in order to compare our results to those in the existing literature which have mostly used a similar fixed-holding period to assess returns.

IV Trading decisions

In this section we investigate the main hypothesis of our study: the sign of the outcomes of past trades influences the current trading volume of the individual investors and, further, the impact of the $Sign$ variable is stronger than that of the $Residual$ variable. We proceed in the following manner. We first examine whether the outcomes of the recent past trades of an investor influence her current trading volume using equation (1)discussed in the previous section, where $RProfit$ is the main independent variable of interest. Next, we decompose $RProfit$ into $Sign$ and $Residual$ using the three-step procedure described in the previous section and estimating equations (2) and (3). Finally, we examine the impact of $Sign$ and $Residual$ on the current trading volume using a suitably modified form of equation (1).

IV.1 *Basic tests*

The dependent variable is $Volume$. The independent variable of interest in equation (1) is $RProfit$. The regression models include three types of control variables: individual-specific controls ($Activity$ and $DProfit$), market controls (Rf and $Index$), and day-of-the week dummies, $Daydummy_1 - Daydummy_4$, to control for the day of the week effect, with Monday as the base day. The models are estimated using unbalanced panel data regression with individual fixed effects and cluster-adjusted standard errors to control for cross-sectional dependence among investors (see Roger, 1993; Petersen 2009; Seascholes and Zhu, 2010). The sample includes 8.8 million observations.

In column (1) of Table 4 below, we report the results of our test of regression model (1). Please note that column report the regression results for the one-lag case, which is our base case. Recall that in this case $RProfit$ includes profits on the immediately preceding day and $DProfit$ includes profits on all other days in the past.We have performed the tests with full as well as winsorized samples (see section III.3 above). Our results are virtually the same for both samples. Hence, in the rest of the paper we report results for the full sample.

(Table 4 here)

From column (1) of panel A, the regression coefficient estimate of $RProfit$ is positive (0.006)and highly significant (t-statistic 28.6). Economically also the effects are significant. For the median investor[6] among the group of investors with positive recent past profits (the winners), a $1 increase in recent past profits results in an increase of $ 1.53 in the investor's current trading volume. For the median investor among the losers, a $ 1 decrease in recent past profits results in a decrease of $ 1.14 in the current trading volume.

The regression coefficient estimates of individual-specific control variables are intuitive. The coefficient of $DProfit$ is positive and has the same size as $RProfit$ (0.006). However, the t-statistic (18.1) is lower, indicating that distant past profits impact current trading decisions less than recent past profits. The result confirms our prediction regarding the wealth effect of past trades discussed in the preceding section. The coefficient of $Activity$ is negative but not significant (t-statistic -1.2). Recall that, by construction, $Activity$ reflects the total number of trades before the current trade. Hence, a larger value of $Activity$ indicates a larger number of past trades. If it takes a large number of transactions to generate a given level of profits (which is controlled for), the

[6]For the median investor in our sample, recent past profits is zero, and therefore the value of $RProfit$, computed as natural logarithm of past profits, is minus infinity. Hence, we consider the median values for the winners and losers separately.

investor is discouraged from increasing trading volume the next time, resulting in a negative association.

Among the market controls, the regression coefficient estimate of $Index$ is positive, but not significant at 1% level. The result is generally consistent with disposition effect. Upward market movement results in more winning stocks and fewer losing stocks than before in an investor's portfolio. The former are sold while the latter are held, increasing trading volume overall. This explanation is also consistent with our result for Rf: the coefficient is negative and significant at 1% level(t-statistic -5.5). As the risk-free rate or time-value of money goes up, stocks generally lose value, resulting in less trading in stocks due again to disposition effect. The regression coefficient estimates of $Daydummies$ are all positive and significant at 1% level. This justifies their inclusion as control variables.

We next use the three-step procedure employing equations (1), (2), and (3) to decompose $RProfit$ into $Sign$ and $Residual$ variables. The results for equation (2) are reported in column (2) of table 4. Recall that in regression model (3) we regress the residual from model (2) on the residual from model (1):

$$e_2 = f\left(e_1\right) + \epsilon_3$$

where e_1 and e_2 are the residuals from models (1) and (2) respectively, and $f(e_1)$ is an appropriate regression function. The exact form of the function is dictated by data. For our sample, the correlation between e_1 and e_2 is observed to be 0.94, justifying a linear regression model. As discussed before, the residual from regression model (3) is our estimate of the $Residual$ variable. The results for equation (3) are not reported to save space.

Next we estimate equation (4) below with $Sign$ and $Residual$ as the main independent variables, but with the same controls as in equation (1), again with individual fixed effects and cluster-adjusted standard errors to control for cross-sectional dependence among investors.

$$Volume_{it_r} = \alpha_i + \beta_1 Sign_{it_r}(k) + \beta_2 Residual_{it_r} + \beta_3 DProfit_{it_r} + \beta_4 Activity_{it_r} + \beta_5 Rf_{t_r} +$$

$$\beta_6 Index_{t_r} + \sum_{g=1}^{4} \beta_{6+g} Daydummy_{t_r g} + \epsilon_{4it_r} \quad (4)$$

The results corresponding to equation (4) for k = 1 are reported in column (3) of panel A, table 4.

From column (3) of panel A, the regression coefficient of $Sign$ variable is positive and highly significant (t-statistic 29.9). Economically also the effect is significant. The current trading volume is 7.7% more for the winners ($Sign$ = 1) than for the losers ($Sign$ = 0)[7]. Note also that the reported t-statistic of $Sign$ far exceeds the corresponding t-statistic of $Residual$ (9.4). In other words, the sign of a past outcome has a more pronounced impact than its size. The results confirm our hypothesis concerning relative impacts of sign and size. The coefficient estimates and the standard errors of the control variables are very similar to those in regression equation (1) discussed above.

As we have noted in the previous section, the error term in regression equation (2) before contains the missing variables included in the error term in equation (1) plus other variables, the relationship between the residuals from the two regressions is in one direction only. Hence, there is no endogeneity due to reverse causation in equation (3). Note also that the coefficients of the control variables corresponding to regression equations (1) and (2) are virtually the same. This indicates that there is no endogeneity problem due to omitted variable bias in model (2), where $Residual$ is the omitted variable.

The adjusted R^2 are in $68\% - 69\%$ range. We also estimate model (1) excluding $RProfit$ from the model (the results are not included in tables but available on request)

[7]For a winner, the increase in $Volume$ is β and the corresponding increase in trading volume is e^{β}. For the loser, the corresponding number is e^0. Hence, the increase in trading volume for winner compared to a loser is $e^{\beta} - 1$.

and observe that the other regression coefficient estimates and their standard errors remain virtually the same. Thus it appears that the impacts of $RProfit$ and of the control variables on the current trading volume are virtually independent.

IV.2 *Robustness tests with different lags*

So far we have considered the results using explanatory variables with one lag: $k = 1$. In order to check the robustness of the results, we test the regression models (1) - (4) with $k = 5$. In this case $RProfit$ includes profits on the immediately preceding 5 days and $DProfit$ includes profits on all other days in the past.The results for equations (1), (2) and (4) with five-lagged variables are presented in panel B of table 4 in columns (4)- (6) respectively. Note that the sample size declines to about 6.6 million, since it includes investors with a minimum of 6 observations

The results for the five-lag case are very similar to the one-lag case. For example, the results for $Sign$ and $Residual$ for the five-lag case, reported in column (4) of table 4, are very similar to the one-lag case. The t-statistics for the two variables are 26.6 and 9.5 respectively, compared to 29.9 and 9.4 in the one-lag case. The results for the control variables are also very similar.

IV.3 *Robustness Tests with probabilistic regression*

In our basic tests discussed above, we studied the association of the actual volume of trades placed by individual investors with the explanatory variables in regression model (4). It is also of interest to examine the probability of changes in trading volume in response to the same explanatory variables. However, in our case it is problematic to conduct a Logit or Probit regression. Not only we have close to 1 million individual fixed effects to deal with, we also control for cross-sectional dependence among investors. Hence, we opt for a linear probability model (LPM) following Wooldridge (2003, p. 455).

For an investor i on trading day t_r, we define y_{it_r} as a binary variable taking a value 1 if $Volume_{it_r} > Volume_{it_{r-1}}$ and 0 otherwise. We use the three-step procedure mentioned earlier to estimate $Residual$. The LPM corresponding to regression model (5) is:

$$y_{it_r} = \alpha_i + \beta_1 Sign_{it_r}(k) + \beta_2 Residual_{it_r} + \beta_3 DProfit_{it_r} + \beta_4 Activity_{it_r} + \beta_5 Rf_{t_r} +$$

$$\beta_6 Index_{t_r} + \sum_{g=1}^{4} \beta_{6+g} Daydummy_{t_r g} + \epsilon_{4it_r} \qquad (6)$$

The test results are reported in table 5 below. Note that the results are obtained for $k = 1$ only, because the object our exercise is to estimate the probability of increase in the trading volume from the investor's previous trading day. The results are duly corrected for heteroscedasticity which is a chronic problem in LPM regressions (see Wooldridge, 2003). We also remove observations that are associated with probability greater than one and less than zero. As a result, the number of observations decline to 7.7 million.

(Table 5 here)

Note that the regression coefficients of the variables $Sign(1)$ and $Residual(1)$ remain positive and highly significant at 1% level. Further, as before, the reported t-statistic of $Sign(17.1)$ exceeds the corresponding t-statistic of $Residual(6.9)$, indicating that $Sign$ has a more pronounced impact than $Residual$ on the probability of an increase in trading volume following a successful trade.

Economically also the results are significant. For the median investor among the winners, each $1 increase in recent past profits leads to an increase of 4.8E(-4) in the probability that the investor's current trades will be larger. For a hundred dollar increase in profits, the probability increases by 4.8%. The corresponding number for the median investor among the losers is 3.7%. From column (2) of table 8, the probability of an increase in current trading volume is 1.7% more for the winners ($Sign$ = 1) than for the losers ($Sign$ = 0).

The LPM results confirm the findings of the basic tests that the individual investors exhibit a strong behavioral bias in their trading activities, and the bias is stronger for the winners than for the losers. The regression coefficients of the control variables are also very similar to what we have observed for the basic tests.

IV.4 *Robustness tests with buy and sell trades separately*

We also estimate regression equations (1), (2), and (4) for buy and sell trades separately. The results (not reported to save space) are very similar to the combined portfolio.

IV.5 *Robustness tests with individual stocks*

Finally, we estimate the regression equations without aggregating the profits across all stocks in an investor's portfolio. In other words, we conduct a stock-by-stock investigation of the role of the $Sign$ variable. The results (not reported to save space but available on request) are very similar to the portfolio approach.

IV.6 *Our results in context*

Our results in this section clearly indicate that the individual investors exhibit behavioral biases in their trading volume. The incremental power of the sign of the past outcomes in explaining current trading decisions is significant and large, and stronger than that of the size of the past outcomes, after allowing for an array of controls, including market controls, observed investor-specific characteristics such as number of trades and past profits, while using individual fixed effects to control for all unobserved cross-sectional variations in individual characteristics and correcting for cross-sectional dependence in individual-level trading. We have seen that the results are robust to various alternative specifications, including different lags, sample sizes, and probabilistic regressions.

Our results are new contributions to the literature. Several existing studies have also noted that the outcomes of past trades influence the trading behavior of individual investors (Barber and Odean 2001; Barber *et al*, 2004; Linnainmaa, 2009). However, the existing studies focus either on the suceess/failure frequency, or on the monetary consequences, of the past trades. They do not consider the sign and size of the past outcomes in one common framework and, consequently, do not observe their relative importance. Analyzing trading records of a sample of households obtained from a discount brokerage house and another sample of investors from a retail brokerage house in the USA, Barber *et al* (2004) conduct a simple binomial test to find that investors exhibit a clear preference for repurchasing stocks that they previously sold at a gain as opposed to repurchasing stocks that they previously sold at a loss. While their test focuses on success/failure of past trades, comparable to our $Sign$ variable, it does not consider size or any other aspects of past trades, or any of the control variables. Using Finnish Central Securities Depository (FCSD) data, Linnainmaa (2009, table IV) also finds a positive relationship between the volume of current trades (dependent variable) and the volume of past trades and returns (independent variables) of individual investors. In a separate regression model, the author finds a positive association between the volume of current trades and the volume of past trades and their success/failure. Though the success/failure variable in the second regression in Linnainmaa's study is similar to the $Sign$ variable in our study, his regression model does not include the other aspects of past outcomes, including their size. Substantively as well as methodologically our work differs from theirs. We isolate the sign and other effects of the outcomes of the past trades, and identify the sign as the main driver of the observed behavioral bias.

Seru *et al* (2009) find that trading experience, computed as the cumulative number of the past trades placed by an investor, has a positive effect on the investor's current return. This result is consistent with our finding that an investor's trading activity, computed in a similar manner, has a negative impact on her current trading volume. Our finding that the regression coefficient of $DProfit$ is positive and significant is also consistent with documented results in the existing literature. Kaustia and Knupfer (2008), find that wealth effects of past trades positively influence current trading decisions. Choi *et al* (2009) find a similar effect for past capital gains.

Using monthly observations on all NYSE/AMEX common stocks during 1962-2002, Statman, Thorley, and Vorknik (2006) find that market-wide trading volume as well as trading volume in individual securities is positively associated with lagged returns. They attribute their findings to investor overconfidence and the disposition effect, respectively. Our findings are not directly comparable to theirs, since our investigations are conducted at the level of individual investors.

V Is sign-influenced trading profitable?

Our results in the previous sections suggest that the outcomes of recent past trades have a positive and significant effect on the current trading decisions of individual investors, indicating overconfidence on the part of the investors. Further, the sign of the past outcomes explains the effect more than the size. Is this type of trading profitable? If so, what we characterize as naive trading driven by a behavioral bias may actually reflect a rational profit-motivated strategy. We examine this issue here. In the first pass, we want to investigate if and how profits from past trades are related to profits from current trades. However, we also want to analyze this issue at a deeper level and conduct further tests to determine what attributes of past profits, sign or size, determines the relationship we observe in our first-pass tests.

For the first-pass test relating profits from past trades to those from current trades, we estimate the following regression model:

$$Profit_{it_r} = \alpha_i + \beta_1 RProfit_{it_r}(k) + \beta_2 Activity_{it_r}(k) + \beta_3 DProfit_{it_r}(k) + \beta_4 Rf_{t_r} +$$

$$\beta_5 Index_{t_r} + \beta_6 Volume_{it_r} + \epsilon_{it_r} \quad (5)$$

The dependent variable is the natural logarithm of profits from trades on trading day t_r for individual investor i. As before, profits are computed over a variable holding period (see section II) above). For robustness check, we also compute profits over a 30-day holding period (see below). The independent variable of interest is $RProfit$, the natural logarithm of the sum of profits from recent past trades. We include $Volume_{it_r}$, volume on the current trading day for investor i, as a control variable. The other control variables are the same as in equation (5), including Rf and $Index$ to control for movements in risk free rate and the market portfolio. The results of the regression are reported in table 6 below.

(Table 6 here)

We focus on the results in panel A of the table ($k = 1$). In column (1) of the panel, recent past profits, represented by $RProfit$, appears to have a strong negative association with current profits: regression coefficient of about -0.19 with t-statistic of -59.9. We find a similar result for $DProfit$ which captures distant past profits (regression coefficient -0.21, with t-statistic -60.2). The results clearly reject persistence in profitability of trades made by an investor over time. Our findings are consistent with Choi *et al* (2009). They find that good portfolio performance in the current period does not predict good performance in the following period.

$Activity$ does not appear to influence current profits (regression coefficient of about -0.02 with t-statistic of -1.2). Note that $Volume$ is positively associated with current day's profit: regression coefficient of about 0.06 (t-statistic of 2.73). This makes intuitive sense. Larger investments lead to larger outcomes, positive or negative. The regression coefficients of other control variables are also intuitive.

As we have indicated above, we want to investigate further the observed negative association between past profits and current profits, and determine if it is due to the sign of past outcomes or their size. To do so, we run the following regression model.

$$Profit_{it_r} = \alpha_i + \lambda_1 Sign_{it_r}(k) + \lambda_2 Residual_{it_r}(k) + \lambda_3 Activity_{it_r} + \lambda_4 DProfit_{it_r}(k)$$

$$+\lambda_5 Rf_{t_r} + \lambda_6 Index_{t_r} + \lambda_7 Volume_{it_r} + \epsilon_{it_r} \tag{6}$$

The variables $Sign$ and $Residual$ are derived from $RProfit$ following the three-step procedure discussed above. The results of the regression are reported in columns (2) and (4) of table 6 for $k = 1$ and $k = 5$ respectively. As before, we focus on the former case. $Sign$ has a pronounced negative effect on current profits: regression coefficient of about -2.02 (t-statistic of -64.2). Economically also the results are significant. Profits from current trades is almost 60% less for the winners ($Sign = 1$) than for the losers ($Sign = 0$)[8]. Interestingly, the variable $Residual$ appears to be positively related to profit from current trades. The regression coefficient is about 1.02 (t-statistic of 28.5). Comparing the regression coefficients of $Sign$ and $Residual$, we see that $Sign$ has a much stronger negative association with current profits than the positive association of $Residual$. In other words, the negative impact of of the success/failure of past trades overwhelms the positive impact of the size of the same trades on current profits. This is an important result and explains the negative association of $Rprofit$ with current profits that we have observed in panel A. Thus, our results not only confirm the findings in the existing literature that past profits and current profits for an individual are negatively associated, but also identify the aspect of the past profits responsible for this negative association.

The regression results for $k = 5$ reported in panel B of table 10 are similar to what we have observed $k = 1$ in panel A.

We conduct a robustness test where the dependent variable is the logarithm of profit over a 30-day horizon instead of a one-day horizon. The results, not reported to save space, are very similar to the results for the one-day holding period.

[8]To determine this, we follow the same procedure indicated in footnote 5 before

VI Importance of Sign

From our results in the previous sections, the sign of past profits appears to explain a major part of the observed impact of past profits on the volume and profitability of current trades for individual investors. We have also noted that the results are robust to a host of alternative specifications.

What proportion of the observed impact of $RProfit$ is explained by $Sign$? In this section we investigate this issue. Following Kruskal(1987), we perform an analysis of the relative importance of the explanatory variables in the two regression equations (4) and (6). The results are presented in table 7, panels A and B.

(Table 7 here)

It is evident from panel A of the table when $k = 1$ that $Sign$ has much more importance than $Residual$ (more than four times) in explaining the current trading volume of an investor. In fact, $Sign$ accounts for 75% of the explanatory power of $RProfit$, while $Residual$ accounts for only 20%. $Sign$ also makes the strongest relative contribution to current trading volume.

From panel B, $Sign$ again has much more importance than $Residual$ (about three times) in explaining the current profit of an investor when $k = 1$. In fact, $Sign$ accounts for 85% of the explanatory power of $RProfit$, while $Residual$ accounts for only 15%. Again $Sign$ has the strongest relative contribution of all regressors.

The results for $k = 5$ are very similar.

VII Alternative explanations

Our results in the previous sections have established that the outcomes of recent past trades have a positive and significant effect on the current trading decisions of individual investors, consistent with what has been noticed in the existing studies on investor overconfidence. Further, the observed effect is largely due to the sign of the past outcomes. Are there other explanations for the results besides investor overconfidence? We examine several plausible alternative hypotheses below. None of them explains the observed patterns of investor behavior satisfactorily.

VII.1 *Learning about trading skills*

Individual investors who have had profitable trades in the past ($Sign$ positive) may think that they have greater skill at stock-picking than others who have experienced losses. Therefore, in their thinking, it is rational for them to increase their trading volume. Those who have experienced losses may behave in an exactly opposite manner. Could the observed positive association between past outcomes and current trading decisions be driven by the investors' learning about their respective investment skills? However, if profitability of past trades is an indicator of superior investing skills, and if losses indicate inferior skills, then we should expect to see persistence in portfolio returns of investors over time. In other words, high (low) portfolio performance in the past should predict high (low) performance in the current trades. However, our results in section V above indicate an exactly opposite pattern (see table 6 above). Past outcomes as well as the sign of past outcomes are associated negatively with current profits. Hence, investors' learning about their investment skills cannot be an explanation for the observed investor behavior.

VII.2 *Private information*

The investors who have private information are likely to have better investment performance and trade more actively than others. If the proportion of informed investors in our sample is high, then this particular behavior pattern could result in a positive association between past profits and current trading volume. The number of informed investors in our sample of individual investors is unlikely to be significant, if any at all. Nevertheless, we investigate this possibility. As a measure of the investment performance of an investor in our sample, we sum the profits on all trades made by the investor over the entire sample period. We do not consider changes in individual portfolio value, because it is meaningful to include only those changes that are brought about by active trading presumably motivated by private information. We then partition the distribution of investment performance estimates into three equal parts: the top third, the middle third, and the bottom third. Private information hypothesis suggests that the investors in the top bracket will exhibit the strongest association between between past returns and current trading, while the investors in the bottom group will show the weakest. To test this implication, we estimate the regression models (1) and (4) augmented with interactions between our main independent variables, $RProfit$ and $Sign$ respectively, and two indicator variables: indicator2 for the medium bracket and indicator3 for the top bracket. Table 8 below reports the results of our investigations for $k = 1$. The results for $k = 5$, not reported, are very similar.

(Table 8 here)

From column (1) of the table, with recent past profits, $RProfit$, as the main independent variable, the investors in the bottom bracket exhibit a strong association between $RProfit$ and the current trading volume (regression coefficient 0.005; t-statistic 21.3). The association for the investors in the medium and top brackets are statistically indistinguishable from the investors in the bottom bracket. Clearly, the results contradict the private information hypothesis. The results with $Sign$ as the main independent variable, reported in column (2) of the table, also refute the hypothesis. The regression coefficient is positive (0.056) and highly significant (t-statistic 22.6) for the bottom bracket, and statistically similar to the results for the medium and top brackets.

VII.3 *Portfolio rebalancing*

Another potential alternative explanation for our results is portfolio rebalancing by the investors in our sample. If the average investor has investments in the risk-free asset in addition to risky stocks, then a positive correlation between the returns of the two asset classes could produce results similar to what we report. Suppose the investor follows a rule of maintaining a fixed dollar amount in the risk-free asset. Then, if the return on the risk-free asset, Rf, goes up, she will move funds from the risk-free asset into her stock portfolio. However, the two returns being positively correlated, it will result in the appearance of higher stock returns leading to a higher volume of stock trading, though in this case portfolio rebalancing is driving the positive association, rather than any behavioral bias on the part of the investors. But note that in this case higher Rf will also be associated more stock trading, which is precisely the opposite of what we find in this study. In fact, in all our tests, Rf is strongly negatively associated with current trading volume (see table 4 before, for example). Therefore, portfolio rebalancing by the investors in our sample does not explain our results.

VII.4 *Wealth effects*

Profitable trades increase an investor's wealth, which may in turn increase her appetite for risk and lead to more investments in risky stocks, inducing a positive association between past profits and current trading volume (Kaustia and Knupfer, 2009). In our tests $DProfit$ represents profits from all past trades until the current phase, and serves as a proxy for changes in wealth. As we have noted, this variable does appear to have a positive impact on current trading decisions. However, controlling for this effect, the impact of our main variables of interest, $RProfit$ and $Sign$, remain very strong (and stronger than that of $DProfit$), as can be checked from tables 4 - 6.

VIII Conclusions

Using a unique and large dataset, in this paper we have contributed several new insights to the existing literature on trading behavior of individual investors.

Decomposing profits/losses of past trades for an investor, represented by $RProfit$, into $Sign$ and $Residual$ variables with the help of a three-step procedure, and using the two variables as separate explanatory variables in our regression tests, has been the most innovative feature of our approach. It has also been the key to our major results. We have seen that $Sign$ has a significant effect on the current trading decisions of individual investors, including volume and frequency of trading. Further, the effect is considerably stronger than that of other aspects of past outcomes including the size of the outcomes. The empirical evidence in section VI above indicates that $Sign$ accounts for 75% of the explanatory power of $RProfit$ when the trading decision is volume of trading. Also, $Sign$ accounts for 85% of the effect of $RProfit$ on the current profits of investors. The results hold after allowing for an array of controls, including observed investor-specific characteristics such as past trading records (including number of trades, gaps in trading activity and trading performance) while using individual fixed effects to control for all unobserved cross-sectional variations in individual characteristics. We have also noted that the economic magnitude of the effects of $Sign$ on current trades is large.

Ours is the first study to consider the relative importance of the sign and size of past outcomes in explaining investor trading behavior. We have seen that the observed behavior results consistently in losses for the investors, suggesting a non-rational motivation. Recent research in experimental psychology has documented that people become less sensitive to the magnitude of a stimulus as the magnitude increases (Hsee and Rottenstreich, 2004; Hsee, Rottenstreich, and Xiao, 2005). Our analysis, based on actual trading data rather than experimental evidence, leads to a similar conclusion. In our setting, as the size of the outcome of a trade increases, it becomes progressively unimportant in the investor's evaluation of the outcome. We also consider several alternative rational explanations for our findings, including trading by investors who are learning about their true investing skills, private information, portfolio rebalancing, and wealth effects of past profits. None of them, we find, explains our results (see section V!! above). As have noted, our findings not only provide new insights into investor behavior, they also throw new light on behavioral patterns documented in the existing literature, such as excessive trading by overconfident individual investors. The findings suggest that the sign of the outcomes of past trades, rather than their dollar size, induces overconfidence in individual traders.

Our findings have important implications for policy-makers and for future research. If individual investors who are mostly small lose from sign-influenced trading, as we document, presumably they could benefit from investor education and guidance. However, any such education is costly, and has been found to be ineffective in tests (Cole and Shastry, 2008). A related policy issue is whether the stock exchanges and agencies in charge of investor protection have a role in providing education and guidance to protect investors from the harmful effects of their own biases, and whether that role is more important in emerging economies. An interesting research issue in this connection is wealth transfer from individual investors to institutional investors. In their study of investor trading behavior in Taiwan, Barber*et al* (2009) find evidence of wealth transfer of the kind and suggest, but do not establish, that the transfer is due to aggressive trading by individual investors driven by a mixture of overconfidence and gambling-like sensation-seeking. The investors in our sample collectively lost 7.66 billion from trading during the sample period (see table 1 above). Our test results in section IX above indicate that the individual investors who are relatively less influenced by the documented behavioral bias are more successful than others. If institutional investors are rational, or at least less biassed than individual investors, then the bias could provide an explanation for the wealth transfer observed in our case. It is a topic for future research.

References

[1] Anscombe, F.J. 1967, "Topics in the investigation of linear relations fitted by the method of least squares (with discussion)", *J.Roy. Statist. Soc., Ser. B*, 29, 1-52.

[2] Arthur, Brian W., 1991, "Designing economic agents that act like human agents: A behavioral approach to bounded rationality", *American Economic Review* 81, 353-359.

[3] Barberis and Thaler, 2003, "A Survey of Behavioral Finance," in George Constantinides, Milton Harris, Rene Stulz eds., *Handbook of the Economics of Finance*, North-Holland.

[4] Brad Barber and Terrence Odean, 1999, "The Courage of Misguided Convictions: The Trading Behavior of Individual Investors" *Financial Analyst Journal*, November/December 1999, 41-55.

[5] Brad Barber and Terrence Odean, 2000, "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors" *Journal of Finance*, Vol. LV, No. 2, April 2000, 773-806.

[6] Brad Barber and Terrence Odean, 2001, "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment" *Quarterly Journal of Economics*, February 2001, Vol. 116, No. 1, 261-292.

[7] Brad Barber, Yi-Tsung Lee, Yu-Jane Liu, and Terrence Odean, 2009, "Just How Much Do Investor Lose from Trade?", *Review of Financial Studies*.

[8] Brad Barber, Terrance Odean, and Michal Strahilevitz, 2004, "Once Burned, Twice Shy: Naïve Learning, Counterfactuals, and the Repurchase of Stocks Previously Sold," Working paper, UC Berkeley.

[9] Bhimasankaram P and D. Sengupta 1996, "The linear zero function approach to linear models" Sankhya, series B, 58, 338-351.

[10] Camerer, Colin F. and Teck-Hua Ho, 1999, "Experience-weighted attraction learning in normal form games", *Econometrica* 67, 827-874.

[11] Charness, Gary, and Dan Levin, 2003, "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect," *American Economic Review* 95, 1300-1309.

[12] Choi, J., D. Laibson, B. Madrian, and A. Metrick, 2008, "Reinforcement Learning and Investor Behavior," forthcoming *Journal of Finance*.

[13] Coval, J. D., D. A. Hirshleifer, and T. Shumway, 2005, "Can Individual Investors Beat the Market?", Working Paper, University of Michigan.

[14] Coval, J. D., and Tyler Shumway, 2005, "Do Behavioral Biases Affect Prices?" *Journal of Finance* 60, 1-34.

[15] Cross, John G., 1973, "A stochastic learning model of economic behavior," *Quarterly Journal of Economics* 87, 239-266.

[16] Erev, Ido, and Alvin E. Roth, 1998, "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review* 88, 848-881.

[17] Gervais, Simon and Terrence Odean, 2001, "Learning to be Overconfident", *Review of Financial Studies*, Spring 2001, Vol. 14, No. 1, pp. 1-27.

[18] Hsee, C. K., Rottenstreich (2004), "Music, Pandas and Muggers: On the Affective Psychology of Value." *Journal of Experimental Psychology*: General. 133, 23-30.

[19] Hsee, C. K., Rottenstreich, Y. Xiao Z. (2005), "When more is better: On the relationship between magnitude and subjective value" . *Current Trend in Psychological Science.* 14, 234-237.

[20] Kaustia and Knupfer, 2008, "Do investors overweight personal experience? Evidence from IPO subscriptions"*Journal of Finance* Vol 63:6 p 2679-2702.

[21] Kruskal, W, 1987, "Relative importance by averaging over orderings," *The American Statistician*, 41, 6-10

[22] Linnainmaa, J., 2009, "Why Do Households Trade So Much?," Working Paper, University of Chicago.

[23] Mosteller, F. and Tukey, J.W. 1977, "*Data Analysis and Linear Regression.*" Reading, Mass. Addison-Wesley.

[24] Odean,Terrance.,1999, "Do Investors Trade Too Much?", *American Economic Review*, Vol. 89, December 1999, 1279-1298.

[25] Petersen, M., 2009, "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches", *Review of Financial Studies*, 22, 435-480.

[26] Rao CR., 1973, "*Linear Statistical Inference and its applications*" Second edition, Wiley

[27] Roth, A. E. and Erev, I. (1995) "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, Vol. 8, No. 1, pp. 164-212

[28] Mark Seasholes and Ning Zu (2010), "Individual investors and Local bias" Fourthcoming *Journal of Finance*

[29] Seru, A., T. Shumway, and N. Stoffman, 2009, "Learning by trading," Forthcoming *The Review of Financial Studies.*

[30] Shawn Cole and Gauri Kartini Shastry, 2009, "Smart Money: The Effect of Education, Cognitive Ability, and Financial Literacy on Financial Market Participation", Harvard Business School Finance Working Paper No. 09-071.

[31] Statman,Meir ., Steven Thorley , Keith Vorkink,2006 "Investor Overconfidence and Trading Volume" *The Review of Financial Studies*, Vol. 19, No. 4 (Winter, 2006), pp. 1531-1565

[32] Wooldridge 2002, "*Econometric Analysis of Cross Section and Panel Data*" MIT press.

Table 1: Trade statistics

The table presents the trade statistics for individual investors from January to June 2006 for Nifty 50 stocks.

Total no. of Investors	0.99mn
Total no. of trades	105.62mn
Total volume traded	INR 1825.67bn
Total no. of trading days	123 days
% of investors traded on at least 2 days	80.97%
% of investors traded on at least 6 days	52.37%
% of winning trading days	51.72%
% of loosing trading days	48.28%
Gains	INR 6.09bn
Losses	INR 7.66bn

Table 2: Variable definitions

We use the subscripts i, t_r, j and l to denote a particular investor, her trading day, stock and transaction number respectively. The subscript r takes values $1, 2, \cdots, N_i$, where N_i denotes the total number of trading days for the i^{th} investor. We use subscript r for t because the trading days are different for different investors. S_{it_rjl} denotes the number of shares traded by the investor i in the l^{th} transaction on day t_r in stock j. P_{it_rjl} denotes the corresponding price. CP_{t_rj} denotes the closing price of stock j on day t_r.

Variable	Description
Individual specific	
$Volume$	$Volume$ is the natural logarithm of the trading volume of an investor on a given trading day. Trading volume of the investor i for the transaction l on trading day t_r in stock j is computed as $S_{it_rjl} \times P_{it_rjl}$. The trading volume of the investor i on trading day t_r is calculated as the sum of the values of all the stocks bought or sold by investor i on trading day t_r.
$Profit$	Profit is calculated by comparing the transaction price to the closing price the day before the next trade.
$Rprofit(k)$	$Rprofit$ refers to recent past profits. $Rprofit_{it_r}(k)$ is computed as the natural logarithm[1] of the sum of profits for preceding k trading days of investor i prior to trading day t_r.
$Dprofit(k)$	$Dprofit$ refers to distant past profits. $Dprofit_{it_r}(k)$ is computed as the natural logarithm of the difference between sum of all profits till the current trading day t_r and $Rprofit_{it_r}(k)$ for investor i. In other words, $Dprofit$ does not include the profits included in $Rprofit$.
$Sign(k)$	Equals 1 if the sum of profits for preceding k trading days of the investor is greater than zero and 0 otherwise.
$Residual(k)$	It is the part of $Rprofit(k)$ impacting the dependent variable over and above the $Sign(k)$.
$Activity$	The number of trades an investor has placed prior to the current trading day.
Market controls	
Rf	The variable Rf represents the average of MIBOR (Mumbai Inter-Bank Offering Rate) and MIBID (Mumbai Inter-Bank InterBid rate) for the current trading day.
$Index$	The variable $Index$ represents the return on the NIFTY index on the preceding trading day.
Day dummies	
$Daydummy_1$	Equals one if the current trading day is Tuesday, and 0 otherwise.
$Daydummy_2$	Equals one if the current trading day is Wednesday, and 0 otherwise.
$Daydummy_3$	Equals one if the current trading day is Thursday, and 0 otherwise.
$Daydummy_4$	Equals one if the current trading day is Friday, and 0 otherwise.

[1] We define natural logarithm of a real number x as

$$\begin{cases} ln(x) & \text{if } x > 0 \\ -ln(-x) & \text{otherwise.} \end{cases}$$

Table 3: Summary statistics

The table presents the basic descriptive statistics of the variables used in our regressions.

Variable	Min	1 Quart	Median	3 Quart	Max	Mean	Std Dev
$Volume$	3.91	9.66	10.75	11.98	22.29	10.87	1.71
$Rprofit(1)$	-17.16	-5.85	2.22	5.67	17.65	0.05	6.02
$Dprofit(1)$	-17.72	-8.28	-6.14	6.55	17.25	-1.92	7.55
$Sign(1)$	0	0	1	1	1	0.52	0.50
$Rprofit(5)$	-17.39	-7.46	-5.19	6.55	17.72	-1.09	7.16
$Dprofit(5)$	-17.72	-8.29	-6.19	6.45	17.23	-2.03	7.52
$Sign(5)$	0	0	0	1	1	0.43	0.50
$Activity$	0.00	0.00	0.00	0.02	286	0.04	1.00

* Calculated for variable holding period.

Table 4: Effect of past outcomes on current trading volume

(Variable holding period)

The table presents the results from estimating the regression equations:

$$Volume_{it_r} = \alpha_i + \beta_1 RProfit_{it_r}(k) + \beta_2 Activity_{it_r} + \beta_3 DProfit_{it_r}(k) + \beta_4 Rf_{t_r} + \beta_5 Index_{t_r} + \sum_{g=1}^{4} \beta_{5+g} Daydummy_{t_r g} + \epsilon_{1it_r}$$

$$Volume_{it_r} = \gamma_i + \theta_1 Sign_{it_r}(k) + \theta_2 Activity_{it_r} + \theta_3 DProfit_{it_r}(k) + \theta_4 Rf_{t_r} + \theta_5 Index_{t_r} + \sum_{g=1}^{4} \theta_{5+g} Daydummy_{t_r g} + \epsilon_{2it_r}$$

$$Volume_{it_r} = \mu_i + \lambda_1 Sign_{it_r}(k) + \lambda_2 Residual_{it_r}(k) + \lambda_3 Activity_{it_r} + \lambda_4 DProfit_{it_r}(k) + \lambda_5 Rf_{t_r} + \lambda_6 Index_{t_r} + \sum_{g=1}^{4} \lambda_{6+g} Daydummy_{t_r g} + \epsilon_{4it_r}$$

Each observation is the aggregation of the data for an investor for a particular day. All the variables used are defined in table 2. The variables *Activity*, *DProfit*, *Rf*, *Index* and *Daydummies* are used as controls. We only consider traders who traded on at least two days during the sample period of 123 days. The regression equation is estimated using unbalanced panel data regression with individual fixed effects. The residuals from the first regression include the effect of the omitted variables, apart from the noise. The residuals from the second regression include the effect of the omitted variables as well as the other effects of *RProfit* excluding the *Sign* in addition to noise. In stage 3, we regress the residuals from second stage on residuals from stage 1 to obtain $Residual$(Results not reported). In the last stage (third equation above), we regress $Volume$ on $Sign$, $Residual$ and other control variables as in the first equation. Columns (1), (2), (3) of panel A correspond to the three equations above respectively for $k = 1$. The results for five-lag case are reported in panel B. The t-statistics reported are adjusted for clustering at the day level.

Independent variables	Panel B (k=5)			Panel A (k=1)		
	(4)	(5)	(6)	(1)	(2)	(3)
RProfit	0.006*** (23.71)			0.006*** (28.61)		
Sign		0.079*** (24.36)	0.073*** (26.59)		0.061*** (29.62)	0.058*** (29.94)
Residual			1.065*** (9.45)			1.028*** (9.43)
DProfit	0.004*** (12.71)	0.003*** (11.54)	0.003*** (11.39)	0.005*** (18.10)	0.005*** (17.29)	0.004*** (17.19)
Activity	-0.004 (-1.60)	-0.004 (-1.60)	-0.004 (-1.61)	-0.003 (-1.21)	-0.003 (-1.20)	-0.003 (-1.21)
Rf	-0.027** (-2.32)	-0.026** (-2.21)	-0.025** (-2.19)	-0.054*** (-5.53)	-0.053*** (-5.47)	-0.053*** (-5.46)
Index	0.000 (0.12)	0.000 (0.13)	0.000 (0.14)	0.001 (0.30)	0.001 (0.35)	0.001 (0.37)
Daydummy1	0.034 (1.48)	0.034 (1.48)	0.034 (1.48)	0.021 (1.17)	0.021 (1.17)	0.021 (1.17)
Daydummy2	0.064*** (3.20)	0.064*** (3.21)	0.064*** (3.21)	0.053*** (3.10)	0.054*** (3.11)	0.054*** (3.12)
Daydummy3	0.065*** (3.26)	0.065*** (3.27)	0.065*** (3.27)	0.049*** (3.00)	0.049*** (3.02)	0.049*** (3.03)
Daydummy4	0.076*** (3.22)	0.076*** (3.22)	0.076*** (3.22)	0.050*** (2.61)	0.051*** (2.62)	0.050*** (2.62)
No. of obs	6645136			8812609		
Adj. $\mathcal{R}^2$	68.99%	68.98%	68.98%	68.92%	68.91%	68.91%

(The figures in parenthesis are t-stats.)

*** -Significant at 1% level ** -Significant at 5% level * -Significant at 10% level

Table 5: Effect of past outcomes on probability of increase in trading volume

The table presents the results from estimating the regression equations:

$$y_{it_r} = \alpha_i + \beta_1 RProfit_{it_r}(k) + \beta_2 Activity_{it_r} + \beta_3 DProfit_{it_r}(k) + \beta_4 Rf_{t_r} + \beta_5 Index_{t_r} +$$

$$\sum_{g=1}^{4} \beta_{5+g} Daydummy_{t_r g} + \epsilon_{1it_r}$$

The specification of this table is similar to that of table 4 except for the dependent variable. For an investor i on trading day t_r, we define y_{it_r} as a binary variable taking a value 1 if $Volume_{it_r} > Volume_{it_{r-1}}$ and 0 otherwise. We estimate the above regression model using linear probability model for the reasons mentioned in the text. We use unbalanced panel data regression with individual fixed effects for estimating the equation.

In the equation below, the variable $RProfit$ is decomposed into $Sign$ and $Residual$:

$$y_{it_r} = \mu_i + \lambda_1 Sign_{it_r}(k) + \lambda_2 Residual_{it_r}(k) + \lambda_3 Activity_{it_r} + \lambda_4 DProfit_{it_r}(k) + \lambda_5 Rf_{t_r} + \lambda_6 Index_{t_r} +$$

$$\sum_{g=1}^{4} \lambda_{6+g} Daydummy_{t_r g} + \epsilon_{4it_r}$$

Column (1) refer to the first equation above and column (2) refer to the second equation.The t-statistics reported are adjusted for clustering at the day level.

Independent variables	Panel A (k=1)	
	(1)	(2)
RProfit	0.002*** (18.21)	
Sign		0.023*** (17.05)
Residual		1.040*** (6.91)
DProfit	-0.000*** (-3.43)	-0.000*** (-5.19)
Activity	-0.004*** (-3.92)	-0.004*** (-3.92)
Market controls	YES	YES
Day dummies	YES	YES
No. of obs	7757709	
Adj. $\mathcal{R}^2$	9.25%	9.29%

(The figures in parenthesis are t-stats.)

*** -Significant at 1% level ** -Significant at 5% level * -Significant at 10% level

Table 6: Impact of Over confidence on current profits

We estimate the regression equation:

$$Profit_{it_r} = \alpha_i + \beta_1 RProfit_{it_r}(k) + \beta_2 Activity_{it_r} + \beta_3 DProfit_{it_r}(k) + \beta_4 Rf_{t_r} + \beta_5 Index_{t_r} + \beta_6 Volume_{it_r} + \epsilon_{it_r}$$

The dependent variable is $Profit$. $Profit_{it_r}$ is the natural logarithm of the profit incurred on a given trading day t_r by investor i. The control variables are as defined in table 2. We use panel data regressions with individual fixed effects to estimate the coefficients.

In the equation below, the variable $RProfit$ is decomposed into $Sign$ and $Residual$:

$$Profit_{it_r} = \mu_i + \lambda_1 Sign_{it_r}(k) + \lambda_2 Residual_{it_r}(k) + \lambda_3 Activity_{it_r} + \lambda_4 DProfit_{it_r}(k) + \lambda_5 Rf_{t_r} + \lambda_6 Index_{t_r} + \lambda_7 Volume_{it_r} + \epsilon_{it_r}$$

Columns (1) and (3) refer to the first equation above and columns (2) and (4) refer to the second equation.The t-statistics reported are adjusted for clustering at the day level.

Independent variables	Panel B(k=5)		Panel A(k=1)	
	(3)	(4)	(1)	(2)
RProfit	-0.155*** (-40.97)		-0.193*** (-59.87)	
Sign		-1.685*** (-46.54)		-2.023*** (-64.17)
Residual		1.121*** (19.46)		1.021*** (28.59)
DProfit	-0.108*** (-38.73)	-0.088*** (-36.63)	-0.206*** (-60.20)	-0.190*** (-60.20)
Volume	0.022 (0.94)	0.011 (0.49)	0.058*** (2.73)	0.048** (2.26)
Activity	-0.000 (-0.01)	0.000 (0.04)	-0.015 (-1.16)	-0.015 (-1.16)
Market controls	YES	YES	YES	YES
Day dummies	NO	NO	NO	NO
No. of obs	6645136		8812609	
Adj. $\mathcal{R}^2$	7.37%	7.31%	9.01%	8.98%

(The figures in parenthesis are t-stats.)

*** -Significant at 1% level ** -Significant at 5% level * -Significant at 10% level

Table 7: Relative importance of $Sign$ and $Residual$

We estimate the regression equations:

$$Y_{it_r} = \alpha_i + \beta_1 RProfit_{it_r}(k) + \beta_2 Activity_{it_r} + \beta_3 DProfit_{it_r}(k) + \beta_4 Rf_{t_r} + \beta_5 Index_{t_r} + \epsilon_{it_r}$$

$$Y_{it_r} = \alpha_i + \beta_1 Sign_{it_r}(k) + \beta_2 Residual_{it_r}(k) + \beta_3 Activity_{it_r} + \beta_4 DProfit_{it_r}(k) + \beta_5 Rf_{t_r} + \beta_6 Index_{t_r} + \epsilon_{it_r}$$

The dependent variable Y in panel A is the natural logarithm of the trading volume on a given trading day. In panel B, it is current profits. The other variables are as defined in Tables 2. This table presents the relative contributions (á la Kruskal(1987)) of various regressors to $\mathcal{R}^2$ after controlling for the individual fixed effects. The results are reported using both one and five-lagged regressors. We use the same panel data regression as in tables 4. Panel A gives the relative contributions using either $RProfit$ or ($Sign$ and $Residual$) along with others regressors towards $Volume$. Panel B gives the relative contributions using either $RProfit$ or ($Sign$ and $Residual$) along with others regressors towards $Profit$.

	Panel A ($Volume$)				Panel B ($Profits$)			
	k=1		k=5		k=1		k=5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rprofit	47.8%		75.0%		59.0%		79.5%	
Sign		36.4%		64.3%		48.9%		59.9%
Residual		9.1%		14.3%		10.9%		21.3%
Dprofit	17.4%	18.2%	12.5%	7.1%	38.4%	37.2%	16.3%	14.0%
Activity	0%	0%	0%	0%	0%	0%	0%	0%
Rf	34.8%	36.4%	12.5%	14.3%	0.2%	0.2%	0%	0%

Table 8: Effect of past outcomes on volume with varying private information

We estimate the regression equation:

$$Volume_{it_r} = \alpha_i + \beta_1 Sign_{it_r}(k) + \beta_{11} Sign_{it_r}(k) * \text{indicator1} + \beta_{12} Sign_{it_r}(k) * \text{indicator2} + \beta_2 Residual_{it_r}(k) +$$

$$\beta_3 Activity_{it_r} + \beta_4 DProfit_{it_r}(k) + \beta_5 Rf_{t_r} + \beta_6 Index_{t_r} + \sum_{g=1}^{4} \beta_{6+g} Daydummy_{t_r g} + \epsilon_{4it_r}$$

The table reports the results for regressions similar to those in table 4 except that the recent past return variables, $RProfit$ and $Sign$ are now interacted with an indicator variable. Investors are first divided into terciles based on performance (measured as total returns in the six month period). For every investor, indicator variables for performance are defined in the following way. If an investor belongs to the highest tercile, the indicator3 gets a value of 1, otherwise 0. If she belongs to the middle tercile, the indicator2 takes a value of 1, otherwise 0. All other variables are as defined in table 2.

Independent variables		
	(1)	(2)
$Rprofit$	0.005***	
	(21.31)	
× indicator2	0.000	
	(0.13)	
× indicator3	0.0003	
	(1.43)	
$Sign$		0.056***
		(22.57)
× indicator2		0.001
		(0.45)
× indicator3		0.003
		(1.33)
$Residual$		1.025***
		(9.10)
$Activity$	-0.003	-0.003
	(-1.21)	(-1.21)
$DProfit$	0.005***	0.004***
	(18.03)	(17.18)
Market controls	YES	YES
Day dummies	YES	YES
No. of obs	8812609	8812609
Adj. $\mathcal{R}^2$	68.92%	68.91%