YALE INTERNATIONAL CENTER FOR FINANCE

Yale ICF Working Paper No. 00-55
August 2001

DISPERSION OF OPINIONS AND STOCK RETURNS: EVIDENCE FROM INDEX FUND INVESTORS

William N. Goetzmann
Yale School of Management
Massimo Massa
INSEAD AND YALE UNIVERSITY

Dispersion of opinion and stock returns: evidence from index fund investors

William N. Goetzmann and Massimo Massa *
Yale School of Management and Insead †

August 2001.

Abstract

We address the issue of how heterogeneity of trade among investors affects stock returns. We develop a model of the dispersion of opinion among investors that has implications for asset pricing. We test the relationship between dispersion of investor opinion and stock returns using a two-year panel of more than 91 thousand individual accounts in a S&P 500 index fund. We show that dispersion of opinion, proxied by the heterogeneity of trade among investor classes, explains part of the returns not accounted for by standard asset pricing factors. We show that the explanatory power of the dispersion of opinion increases at the very time when standard pricing models based on standard asset pricing factors fare worse.

JEL classification: G11,G12,G14.

Keywords: Learning, asset pricing, market confidence, behavioral finance.

---

*We thank Fidelity for providing us with the data for this study. We thank the International Center for Finance at the Yale School of Management for providing research support.

†Corresponding author: Massimo Massa, Finance Department, INSEAD, Boulevard de Constance 77305, Fontainebleau Cedex, France. e-mail: massimo.massa@insead.fr. Fax : (33)1 60 72 40 45.

1 Introduction

Differences of opinion among investors are widely recognized as potential determinants of asset prices. While liquidity motives may explain much of the daily trading volume in asset prices, the actions of informed traders and traders who believe they are informed play a recognized role in models of financial markets. Theoretical analysis has extensively studied the impact of differences of opinion on price. In this paper we empirically address this issue by constructing a factor that correlates to the dispersion in investor opinion and showing that it provides additional explanatory power in a standard asset pricing model.

We test the model with a panel of investor accounts – a panel for which we can essentially eliminate private information as a motive for trade. All of our investors hold shares in an S&P 500 Index fund. In our tests we develop time-series factors of trading activity based strictly on heterogeneity among investors, not on asset prices. When we add these factors to the standard asset pricing framework, we find that they are strongly significant. We further find that, in periods when the standard asset pricing factors perform poorly, our heterogeneity-based factors perform well. While our results can be interpreted as evidence that the behavior of uninformed investors affects stock price dynamics, it is also consistent with a framework of rational, yet heterogeneous, reaction to market-wide information.

The finance literature has extensively analyzed the link between differences of opinion among investors and asset prices. Differences of opinion have been justified in terms of either heterogeneity of beliefs or asymmetry of information. Williams (1977) incorporates heterogeneity of opinion in the standard CAPM framework and shows how it affects market returns. DeTemple and Murthy (1994) prove that the equilibrium interest rate itself is a function of investors' opinions, weighted according to the fraction of total wealth they hold. Kraus and Smith (1989) argue that, even in the absence of new information about security payoffs, a change in opinion may move prices. The fact itself that investors are imperfectly informed about each other's endowments creates and reinforces uncertainty and preserves heterogeneity of opinion at equilibrium. This "market created risk" affects prices and equilibrium levels of returns.[1] Harris and Raviv (1993) develop empirical implications about

[1]Also, Kim and Verrecchia (1991), Grundy and McNichols (1989) and Shalen (1993), He and Wang (1993) and Biais and Bossaerts (1998), by explicitly model the link heterogeneity of beliefs to trading volume and volatility, identify a positive direct relationship between dispersion of beliefs and both volume and price volatility.

volume, volatility and price dynamics due to degrees of heterogeneity in the way investors interpret news.

While in all these models investors' heterogeneity plays a key role in the price-formation process, the restrictions directly testable in terms of investors' trade and holding positions are very few. Only recently Wang (1993) and He and Wang (1993) explicitly relate the heterogeneity of trade among investors and asset prices to asymmetry of information.

To date, however, the empirical support for these theories has suffered from the lack of a good proxy. Aggregate market trading volume and open interest have been identified with the dispersion of investor opinion. However, measures relying on aggregate data do not directly capture investors' differences of opinion. Heterogeneity of investors' trade should intuitively be the best candidate. However, there are two main obstacles to its use.

The first is the fact that the link between heterogeneity of trade and asset returns is not directly evident. That is, it is not clear how the process of aggregating investors' heterogenous holdings may affect the determination of stock returns. Indeed, Lo and Wang (2000) and Bossaerts, Plott and Zame (2000) have recently come to opposite conclusions regarding the relationship between investors' holdings and asset prices. Lo and Wang suggest that a direct relationship between trading volume and stock returns exists and can be estimated on the basis of the restrictions on investors' holdings that are implicit in the CAPM framework. Bossaerts, Plott and Zame, on the contrary, use repeated experiments to show that CAPM pricing relationship can hold regardless of the way holdings are distributed among investors. That is, the aggregation results implicit in CAPM do not seem to place any restriction on investors' individual holdings.

A second problem is due to the fact that, while many of the theoretical models cited above have been motivated by the compelling empirical evidence of temporal regularities in asset price patterns and *overall* trading, direct evidence on *individual* investor behavior has been illusive. Studies by Grinblatt, Titman and Wermers (1995) and Lakonishok, Shleifer, Thaler and Vishny (1991), Edelen and Warner (1999) and Goetzmann and Massa (2000), focus on the behavior of institutional managers, as opposed to individuals.[2] Information about individual investor behavior has been very difficult to obtain. Schlarbaum, Lewellen and Lease (1978),

[2] Also, Gompers and Metrick (1998) study the equity holdings of large institutions for its implications for liquidity.

Lakonishok and Maberly (1990) and more recently Grinblatt and Kellaharju (1999) use individual investor account data in order to analyze how investors (or investment groups) trade in individual securities. In particular, Grinblatt and Kellaharju (1999) report that the actions of foreign investors alone significantly correlate to price changes in the most active stocks in Finland, suggesting that foreign investors may be the salient group. However none has directly studied how *heterogeneity of investors' trading patterns affects stock returns.*

In this paper, we focus on the potential asset pricing role played by heterogeneity of investor opinion. We explicitly incorporate the possibility of irrational or biased beliefs about the market into an asset pricing model. In a manner similar to DeLong, Shleifer, Summers and Waldmann (1990), the behavior of the biased investors, while irrational, has important pricing implications. Investors' learning errors are a source of risk, and their impact on prices is compounded by the degree of investors' irrationality.

In the empirical analysis, we use information on individual investors' purchases and sales of index fund to construct measures of dispersion of opinion. In particular, we consider a panel of more than 91,000 investor accounts in an S&P 500 Index Fund over a two year horizon. This allows us to explicitly test the relationship between investors' heterogeneity of trade, dispersion of opinion and asset returns.

We focus on index fund investors because their behavior provides a reasonable proxy of overall market. Indeed, the fact that index fund investors explicitly choose an index fund as opposed to a managed fund allows us to focus on the relationship between stock prices and dispersion of opinion about the stock market as a whole, rather than the relative investment prospects for individual securities, as most of the literature has hitherto done. Furthermore, given that index fund investors are more likely to invest for the long run, we can use their decision to enter the market as a sign of investors' confidence in long-term market outlook. While we expect much of the trade in the fund is due to liquidity needs, we expect that at least some of the correlated trade is driven by bullish or bearish beliefs, and thus may provide a noisy instrument of dispersion in market sentiment at each point in time.

Our econometric analysis provides some support for the theory, and suggests that dispersion of opinion may explain part of the variance in returns not accounted for by the standard asset pricing factors. An alternative explanation is that returns and dispersion of opinion are both influenced by unidentified economic factors.

The paper is structured as follows. In Section 2 we lay out the model and its testable restrictions. In Section 3 and 4 we describe the data and the way we use it to construct a micro-based index of dispersion of opinion. In Section 5 we report the empirical tests. A brief conclusion follows.

2 The model

The economy.

The model is a simplified version of the standard dynamic rational expectations model that accounts for investors' irrationality. We use it to define the main testable restrictions. In particular, we build on Wang (1993) and Hau (1998). We consider an exchange economy with a single good. Dividends (D) are a function of the output process (Π) of the economy. Dividends and output follow respectively:

$$dD = \Pi dt + \mathbf{b}_D d\mathbf{z}, \text{ and } d\Pi = a_\Pi(\overline{\Pi} - \Pi)dt + \mathbf{b}_\Pi d\mathbf{z}, \tag{1}$$

where $\mathbf{b}_D = (\sigma_D, 0)$, $\mathbf{b}_\Pi = (0, \sigma_\Pi)$ and $\mathbf{z}$ is a vector of stochastic processes.

One stock is traded, with a price P. It is in positive net supply, with a stationary supply level, in the long run, normalized to 1. In line with the standard literature (Sundaresan 1983, Wang 1993, He and Wang 1993, Naik 1997) we assume a constant interest rate (r).

The information structure.

We assume two classes of investors: sophisticated investors and unsophisticated investors. There are ω sophisticated investors and $(1-\omega)$ unsophisticated investors.[3] The sophisticated investors have access to a privileged source of information on dividends and output. Newsletters, brokers' reports, specialized press and research in general are examples of this type of information. For simplicity and we no loss of generality, we assume them to be fully informed. The unsophisticated investors, instead, do not have access to any privileged source of information and learn D and Π only by observing prices. That is, $\mathcal{F}_t^i = \{D, \Pi\}$ and $\mathcal{F}_t^u = \{P\}$. Let's define the vector of the state variables as $\mathbf{m} = (D, \Pi)$.

Theorem 1. Investor learning

The estimated state variables $(\widehat{D}, \widehat{\Pi})$ *for the jth agent are derived applying an*

[3]This grouping of investors is meant to follow the standard definition used by regulatory authority in the AngloSaxon countries.

optimal filtering leading to the following system of equations:

$$\begin{bmatrix} d\widehat{D} \\ d\widehat{\Pi} \end{bmatrix} = \begin{bmatrix} \widehat{\Pi} \\ a_{\pi}(\overline{\overline{\Pi}} - \widehat{\Pi}) \end{bmatrix} dt + \mathbf{H}d(\widehat{\mathbf{z}})$$

(see Wang, 1993).

The information variance-covariance matrix ($\mathbf{H}$) and the new "induced" stochastic processes ($d\hat{\mathbf{z}}$) are defined by the parameters of the learning process.[4] The learning error of the investors make is $\Delta = \widehat{\Pi} - \Pi$.[5] It follows:

$$d\Delta = -a_{\Delta}\Delta dt + b_{\Delta}dz. \tag{2}$$

As investors learn, they update their estimate as well as the conditional variance of their forecasts (Square Forecasting Error or SFE), where $SFE = E[(\widehat{\Pi} - \Pi)(\widehat{\Pi} - \Pi)']$.

We assume that the unsophisticated investors may make systematic learning errors. In particular, unsophisticated investors may keep unaltered the "degree of confidence" in the economy, regardless of their market observations. In practice this means that they do not update the conditional variance of their estimates (SFE). We therefore consider the equilibrium where learning is rational and compare it to other equilibria where investors do not optimally update the SFEs.

These systematic errors may be interpreted as deviations from the optimal investment rule due to some social or behavioral motives (Daniel, Hirshleifer and Subrahmanyam, 1997, ODean, 1998). Therefore, the less sophisticated investors in part behave as the noise traders described by DeLong, Shleifer, Summers and Waldmann (1990). They "falsely believe that they have special information about the future price of the risky asset. The may get their pseudo-signals from technical analysts, stockbrokers, or economic consultants and irrationally believe that these signals carry information. Or in formulating their investment strategies, they

[4]The signal is: $s = p_D^* D + p_{\Pi}\Pi$. The estimated state variables and the signal follow: $d(\mathbf{m}) = [\mathbf{a}_{m0} + \mathbf{a}_{mm}\mathbf{m}]dt + \mathbf{b}_m d\mathbf{z}$ and $d(s) = [\mathbf{a}_{s0} + \mathbf{a}_{ss}s]dt + \mathbf{b}_s d\mathbf{z}$. Let's define $\mathbf{q}_{mm} = \mathbf{b}_s\mathbf{b}_s'$, $\mathbf{q}_{ss} = \mathbf{b}_s\mathbf{b}_s'$, $\mathbf{q}_{ms} = \mathbf{b}_m\mathbf{b}_s'$,while $\mathbf{y}$ is the matrix of the conditional square forecasting errors for the state varables equal: $\mathbf{y} = E[(\widehat{\mathbf{z}} - \mathbf{z})(\widehat{\mathbf{z}} - \mathbf{z})']$, .where the first entry of $\mathbf{y}$ is $SFE = E[(\widehat{\Pi} - \Pi)(\widehat{\Pi} - \Pi)']$. Then, we find that $\mathbf{H} = (\mathbf{y}\mathbf{a}_{sm}' + \mathbf{q}_{zs})\mathbf{q}_{ss}^{-1}$,and $h_1$is the first entry of H. At the stationary learning, $\mathbf{y}$ can be derived as the steady state solution of the Riccati equation: $\dot{\mathbf{y}} = \mathbf{a}_{m,m}\mathbf{y} + \mathbf{y}\mathbf{a}_{m,m}' + \mathbf{q}_{m,m} - \mathbf{H}\mathbf{q}_{m,m}\mathbf{H}' = 0$, (Fleming and Ryshel, 1976). The values of $\mathbf{y}$ and $\mathbf{H}$ define the solution at the equilibrium characterized by rational learning.

[5]Investors actually make errors about both state variables (i.e., $\Delta_{\Pi} = \widehat{\Pi} - \Pi$ and $\Delta_D = \widehat{D} - D$). However, it can be shown that the equilibrium is a function of just one learning error (Wang, 1993). We consider $\Delta = \Delta_{\Pi}$. In particular, $a_{\Delta} = -a_{\Pi} + h_1(p_{\Pi}a_{\Pi} - p_D^*)$, $b_{\Delta} = p_D^* h_1 b_D + b_{\Pi}(h_1 p_{\Pi} - 1)$.

may exhibit the fallacy of excessive subjective certainty that has been repeatedly demonstrated in experimental contexts since Alpert and Raiffa (1982)".

This allows us to see how the deviation from rationality of the unsophisticated investors affects equilibria. The distance of SFE from the rational one captures in some way investors' "degree of overconfidence". If $SFE > SFE^e$, that is unsophisticated investors' conditional variance is greater than the rational one, unsophisticated investors perceive assets riskier than they actually are. Conversely, if $SFE < SFE^e$, that is unsophisticated investors' conditional variance is lower than the rational one, unsophisticated investors perceive assets less risky than they actually are. The first case would correspond to underconfidence and the second case to overconfidence. Therefore, this approach shares part of the salient features of the standard rational expectations models based on heterogeneous information (Wang, 1993, He and Wang, 1993) and part of the features of the behavioral models based on investor "irrationality" (DeLong, Shleifer, Summers and Waldmann, 1990).

Investor preferences

Both classes of investors rationally maximize long term profits (Φ_t) by solving the following problem:

$$\begin{aligned} MaxE \int_{s=t}^{\infty} e^{-r(s-t)} \left[d\Phi_s - \frac{1}{2}\rho(d\Phi_s)^2 \right] \qquad (3) \\ s.t. \quad : \quad d\Phi_s = X_t[dP_t + (D_t - rP_t)dt] = X_t Q_t dt, \end{aligned}$$

where X_t is the amount invested in stocks and ρ is the degree of risk aversion. Q_t represents the excess return over the riskless asset. It can be represented as the return on a zero-wealth portfolio long one share of stock and fully financed by borrowing at the risk-free rate. To focus on the informational effects, we assume that all the investors have the same degree of risk aversion.

Theorem 2. Investor demand

The demand for stock of the sophisticated and unsophisticated investors are, respectively,

$$X_t^i = \frac{E[dR_t|F_t^i]}{\rho E[dR_t^2|F_t^i]} = \frac{e_0 + e_\Delta \Delta_t}{\rho V^u} \text{ and } X_t^u = \frac{E[dR_t|F_t^u]}{\rho E[dR_t^2|F_t^u]} = \frac{e_0 + e_{\widehat{\Psi}} \widehat{\Psi}_t}{\rho V^u}, \qquad (4)$$

where $e_0 = -rp_0$, $e_\Delta = -(r - a_\Delta)$ *and* $V^u = (1 + p_\Delta h)^2$ *(see Wang 1993 and Hau 1998).*[6]

[6]To solve the model we follow the standard approach. We first conjecture a linear pricing

Theorem 3. Equilibrium prices.

There exists a stationary rational expectation equilibrium, where the prices of the assets are a linear function of the underlying state variables.

$$P_t = (\phi + p_0) + p_D^* D_t + p_\Pi \Pi_t + p_\Delta \widehat{\Pi}_t = \Phi_t + p_0 + p_\Delta \Delta_t \tag{5}$$

where $\Phi_t = E_t[\int_t^\infty e^{rs} D(s) ds] = \phi + p_D^* D_t + p_\Pi^* \Pi$, where: $\phi = \frac{a_\pi p_\pi^* \bar{\pi}}{r}$; $p_D^* = \frac{1}{r}$; $p_\pi^* = \frac{p_D^*}{r + a_\pi}$ *(see Wang 1993).* Φ_t captures the fundamental value of the stock as defined in terms of net present value of future dividends. It corresponds to the value of the stock in the case of perfect information ($\Delta_t = 0$). This would be the case if unsophisticated investors were perfectly informed and were not affected by the climate of confidence.

2.1 Testable restrictions

By using equations 4 and 5, after some manipulation, we find the relationship between stock returns and investor demand. In particular, rewriting equation 5 in terms of investor demand schedules we have:

$$P_t = \alpha + \Phi_t + \beta(X_t^i - X_t^u) = \alpha + \Phi_t + \beta Disp_t, \tag{6}$$

where $\alpha = p_0 + \phi + p_\Delta \frac{e_0(1-\omega)}{\omega e_\Delta \rho V^u}$, $\Phi_t = p_D^* D_t + p_\Pi^* \Pi_t$ and $\beta = \frac{(1-\omega e_0)}{\omega e_0} p_\Delta$. The variable $Disp_t = [(1-\omega)X_t^u - \omega X_t^i]$ is a proxy of the dispersion of the beliefs of the investors constructed on the basis of their holdings. Therefore, stock returns follow:

$$d(P_t) = d\Phi_t + \beta dDisp_t, \tag{7}$$

The first testable restriction is a direct relationship between heterogeneity of investor trade ($Disp_t$) and asset returns. From now on, we will define the sources of uncertainty due to the fundamentals (Φ_t) as "standard asset pricing factors" and the sources of uncertainty due to investors' learning and differences of opinion ($Disp_t$) as "dispersion of opinion-related factors".

If the informational content of the dispersion of opinion-related factors is orthogonal to the one contained in price-based information, data on the purchases and sales

function. We then solve the investor's inference and optimization problems. Finally the market clearing condition allows us to back out the value of the coefficients for the linear pricing form we have conjectured. The solution is along the lines of Wang (1993) and Hau (1998), modified to account for quadratic utility function, no noise trade and unknown dividends.

of different types of investors becomes useful to forecast stock prices.[7] This delivers us a testable restriction: in equation 7

$$H1_0 : \beta = 0. \text{ and } H1_a : \beta \neq 0. \tag{8}$$

Furthermore, the loadings on the standard asset pricing factors are not affected by over(under)confidence in the market, investors' learning errors or information asymmetry. Conversely the loadings on the dispersion of opinion (i.e. β) depends on both the relative size of each class of investors (ω) and the degree of under(over)confidence ($SFE - SFE^e$). Let's examine this in more detail. In Figure 1 we report simulated values of β for different fractions of sophisticated investors in the market (ω) and different levels of $SFEs$. [8] At the "rational" equilibrium, that is at the equilibrium where investors optimally update their opinions, $SFE = 0.0024$.

It appears that β is a direct function of both overall information asymmetry and the degree of investor under-confidence ($SFE - 0.0024$) [9]. Overall information asymmetry is a function of both the informational uncertainty of the less sophisticated investors and the ratio between sophisticated and unsophisticated investors. The impact of overall information asymmetry on β increases as ω approaches either 0 or 1 and is minimum for $\omega = 0.5$. Indeed, if there are very few sophisticated investors (ω close to zero), the induced uncertainty (Δ_t) is high and compounds the effects due to dispersion of opinion. On the other hand, if the unsophisticated investors are very few (ω close to 1), the strong asymmetry of information compounds the impact of dispersion of opinion. The same degree of difference in opinion between investors will induce the less sophisticated investors to require a greater risk premium as they are conscious of being surrounded by more sophisticated investors.[10]

The role of informational asymmetry is itself compounded by the degree of risk aversion. The more investors are risk averse (greater ρ), the more dispersion of opinion impacts prices (i.e., β is greater). This can be seen by comparing the three graphs in Figure 1, based on risk aversions respectively equal to 5, 10 and 20.

[7] This last one being a better proxy of fundamentals-related factors.

[8] We consider "plausible" values of b_D and b_Π, that is the ones in general estimated and assumed in the literature (Campbell and Kyle, 1993, Veronesi, 1999).

[9] Indeed, dividends are the signals that uninformed investors observe to filter out the value of Ψ_t.

[10] Wang (1993) shows that if the fraction of informed investors is very high, this increases the chance for the less informed investors of dealing with someone more informed than them. The worsening of the informational disadvantage of the less informed investors increases the premium they ask for informational asymmetry.

Furthermore, the impact of the dispersion of opinion-related factors on stock returns is a direct (positive) function of the degree of under-confidence. The more the investors make unjustified gloomy assumptions about the state of the economy (i.e., $SFE > 0.0024$), the stronger the impact of the dispersion of opinion on stock returns.

This delivers us the second testable restriction: we can identify two "regimes" characterized in terms of market information asymmetry/investors' degree of under-confidence. The first one is characterized by high asymmetry and high under-confidence (High regime). The second by low asymmetry and low under-confidence (Low regime). The testable restriction is that:

$$H2_0 : |\beta_H| = |\beta_L|, \text{and } H2_a : |\beta_H| > |\beta_L|, \tag{9}$$

where β_H and β_L are the impact of the dispersion of opinion on prices in the high regime and in the low regime respectively. We will now proceed to test these implications.

It is interesting to note that "irrationality" by itself does not affect returns, but merely compounds the effects generated by the dispersion of opinion and makes them more severe. Indeed, in the limiting case where all investors are equally informed (or ignorant), irrationality does not affect stock returns.[11]

3 The data

Fidelity provided us with anonymous individual account activity in the Fidelity Spartan Market Index Fund over the years 1997 and 1998. The objective of the fund is to closely match the returns to the S&P 500 Index while keeping management fees, transactions costs and other expenses to a minimum. Over the five years ending in 1999, the fund returned 27.51% per year compared to the S&P 500's return over the period of 27.87%. The fund has a short-term trading fee of 1/2 % for redemptions that occur within 90 days, a minimum initial investment of $10,000 and a minimum required balance of $5,000. These minimums are less for a retirement account. The two years of our study were both banner years for the S&P 500. It grew by 33% in 1997 and by 28.5% in 1998. The fund also grew dramatically over the two-year period – from $1,597.5 million at the end of 1996 to $7,149.9 million at the end of

[11] Indeed, in such a case, investor irrationality would only affect α - i.e. the level of prices - and would not vary stochastically over time.

1998 growing by a factor of two, after the effect of the growth in share prices is taken into account.

We have daily activity records for all accounts that existed or were formed during the two-year sample period. All individual identifying characteristics of these accounts were removed. The accounts are only identified by type which we sorted into four general categories: Individual, Tax-Benefited, Fiduciary and Trust or Group. Table 1 describes our sample. After screening for various data errors (such as accounts with withdrawals that exceed balances) we have a total of 90,768 accounts. We have 259,616 transactions of which 83% are purchases of shares and 17% are share redemptions. The largest category of investor (66,903) is the Tax-Benefited account – principally IRA and Keogh plans. Next is individual account (16,185). We have a small number of Fiduciary accounts (5,493) which include Executors, Guardianship and Trusts. The Group category (2,179) includes Investment Clubs, Partnerships and other accounts that are held in the name of an association of some sort.

How big are the investor accounts? Because accounts begin and end within the sample, determining an appropriate scale measure for the typical account is not trivial. We calculate the average running balance [RB] by taking the average number of shares held by an investor over the period for which the account is open. The average individual RB is 400 shares, or about $28,000 to $36,000, with the median individual account at less than half that. As a measure of activity in the account, we calculate turnover ratio [T] as the absolute sum of the number of share purchases and sales divided by the running balance. Thus, a perfectly passive investor who had 100 shares at the beginning of the period and held them through the end would have a turnover ratio of one. In Table 1, the median turnover ratio for all accounts is slightly greater than one. However, the mean is dramatically higher suggesting that some accounts have a lot of activity.

Table 1 also reports an Investor Profit Ratio. This is a measure of investor profits due to the timing of their flows in and out of the fund. It is not the standard time-weighted rate of return typically used to measure portfolio performance. The time-weighted rate of return would simply equal the return to the index fund over the period of the investor account's existence and would be unaffected by how much money was in the account at different times. As such, it would not provide a measure of timing skill relative to a meaningful alternative. Instead, we use a standard accrual method for profit calculation. The capital appreciation of each share purchased is tracked separately for the investor, and profits are defined as the accumulated growth in all share values at the termination of the account or the end of the sample period. This profit is scaled by the capitalization of the net value of share purchases and sales invested at the beginning of the sample period. In effect, we report timing profits by comparison to a benchmark buy and hold strategy, where we assume the investor could have placed all of his or her money in the fund at the beginning of the two-year period, as opposed to distributing the contributions throughout the period. [12]

4 Construction of an index of dispersion of opinion

4.1 Alternative definitions of dispersions of opinion

Two variables have been traditionally used in the literature as proxies for dispersion of opinion: open interest in options (indicative of agents possibly agreeing to disagree about the prospects of the underlying security) and overall volume of trade. The problem with the aggregated flows, however, is that it is not possible to identify the behavior of different classes of investors. Therefore, by definition, they provide a very poor proxy of investors' heterogeneity. Our dataset, by allowing us to separate the inflows and outflows by investor classes, provides us with an opportunity to calculate a better measure of dispersion of opinion.

As a first rough measure of dispersion of opinion we use the absolute value of the difference between the trades (purchases and sales) of different classes of investors. In particular, if we consider N classes of investors, we can construct the following

[12] This is an imperfect measure, since it relies on certain assumptions that may be unrealistic. Among these assumptions is that the investor has the money to buy shares at the beginning of the sample period, rather than when shares were actually purchased. What we attribute to strategic delay in investment may simply be investor illiquidity. Because of this issue, we also considered scaling terminal share values by the gains to a dollar-cost-averaging strategy that effectively distributed net share purchases equally through the sample period. This however would not change the relative rankings of investors, but only the absolute value of the Profit Ratio. The second major limitation of the investor profit ratio is that many of the accounts in our database opened after the beginning of our sample period. Incoming investors may simply have switched from another S&P index fund rather than cash. Given the high return to the S&P in 1997, latecomers to the fund will typically have a low profit ratio. Because the profit patio measure has limitations, we make no claim that it perfectly measures relative investment skill. It is reported simply to describe sample distributional characteristics and not as an indication of skill across account type.

metric of investors' dispersion of trade:

$$Disp_t = \sum_{j=1}^{N} \sum_{j=1}^{N} Abs\left(X_t^i - X_t^j\right) \text{ for } j = 1, ...i, ...N,$$

where X_t^j is the holding of the jth investor at time t. $Disp_t$ postulates a direct relationship between dispersion of holdings of the investors and dispersion of opinion.

4.2 Alternative ways of identifying and grouping investors

How do we select the N classes of investors whose aggregate trades make up our X_t^j? The selection is done by grouping the individual accounts on the basis of the trading characteristics of the investors. However, this characterization of the investors is not without problems, as it can potentially suffer from an endogeneity bias, being the classification based on in-sample data. We deal with this problem considering two alternative ways of grouping the investors.

4.2.1 A first set of groupings

As a first approach, we consider several groupings constructed by using different criteria with different exposure to the endogeneity bias. We then compare the results across specifications to see whether they are robust to the change of the specification. If a specification based on a selection criterion orthogonal to the tests we are carrying out delivers results consistent with the ones of a specification more subject to endogeneity, we can safely assume that the endogeneity error is not very significant.

In particular, we identify investors on the basis of five criteria: the amount of money invested in the index fund on average (Average Holdings), the money they have invested at the end of the period (Running Balance), the dispersion of the holdings over time (Holding Dispersion), their trading frequency (Number of Transactions) and the rotation of their portfolio (Turnover). Average Holdings are defined as the number of shares the investor has in the fund multiplied by the length of time they are held, the Dispersion of Holdings is the standard deviation of the holdings over time. Turnover is calculated as the absolute sum of purchases and sales in the fund divided by the average running balance. Running Balance is constructed as the average holdings standardized by the amount of time they are held. Investors are then ranked in 50 groups in ascending order.

4.2.2 Groupings based on "consistent" investors

Alternatively, we group investors in sample and then trace their behavior out of sample. In order to do this we define investors on the basis of their conditional pattern of share purchases and redemptions. That is, in terms of the way they react to past return and volatility.

Given the type of investors - mostly investors with a long-term investment horizon - we cannot hope to capture and analyze investors' short-term trading strategies. However, we can use investors' reactions to returns and risk as a crude way to identify investors who display some consistent pattern. Our aim is to identify them in one period and follow their behavior out-of-sample. This should overcome any possible endogeneity bias. A test of consistency of behavior over time is then performed in order to check for the robustness and economic meaningfulness of our classification and to be sure that we are actually capturing some behavioral characteristics and not some statistical fluke. [13]

In particular, we define as positive feedback traders investors who purchase when the market rises and sell when the market falls in the previous trading session. Negative feedback traders, on the contrary, buy after a drop in the market and sell after a rise. We also classify them in terms of their response to changes in the implied volatility of the S&P 500.

Our classification of investors as positive and negative feedback traders is based on a binomial test of the differences in proportions applied to daily investor purchases and sales and the daily market return. We define as a positive feedback trader an investor whose frequency of share purchases following days after a market rise is greater than would be expected given a random distribution of share purchases of the same number within the sample period. A negative feedback trader, is defined analogously as an investor who sells shares conditional upon an increase in the market on the previous day, and buys conditional on a market downturn. The null hypothesis for both types is that the ratio of purchase-days to non-purchase-days, conditional upon previous day's market direction, is equal to the unconditional ratio

[13] It would be possible to define positive and negative feedback trading over much longer horizons and in many other alternative ways. Indeed, for studies of momentum investing, for example, it would be useful to condition behavior on the market performance over previous weeks, months or years. For instance, Grinblatt and Kellaharju (1998) base their analysis upon the past several months as opposed to days. Our choice of the daily horizon is based upon previous analysis of aggregate index fund flows (Goetzmann and Massa, 2000.), where some evidence is found that, on average, index fund investors reacted negatively to the previous day's market drop.

of up (or down) days for the market. Since investors trade relatively infrequently in our sample, we cannot employ the normal approximation to the binomial and thus, critical values for rejection of the null are given by summation of the binomial frequencies up to a probability level less than the critical value of 10% for a one-sided test. We apply this test to each investor's inflows and outflows separately. The same procedure is used to classify investors according to the change in implied volatility in the preceding trading day. [14]

Table 2 reports the classification of accounts according to whether they have positive or negative feedback tendencies. The top panel reports results for all accounts and the bottom panel restricts the analysis to accounts with eight or more transactions in the period. The distribution for inflows and outflows into individual accounts suggests that the negative feedback investors are slightly more common than the positive feedback investors. Almost 25% of the accounts display a negative-feedback trading tendency, while only 12% display positive feedback characteristics. This is true across all four categories of accounts and is consistent with Grinblatt and Kellaharju's findings that contrarians are more common in their sample than momentum investors. [15] Accounts with more than eight transactions show a different tendency from the general population, displaying some tendency towards negative feedback. The other three groups appear to strongly favor positive feedback – on balance more than 50% of the frequent traders appear to be positive feedback investors, vs. 37%. [16]

In order to test for the consistency of the behavior, we examine whether investors identified as feedback investors in the first period are more likely to be feedback traders investors in the second period. We again use an odds-ratio test based on a two-by-two table, considering all accounts that existed over two subperiods: 1/1/1997 to 31/12/1997 and 1/1/1998 to 31/12/1998. For each period we use the proportion statistic described in the preceding section to identify investors as positively or negatively reacting to either returns or volatility, where the median

[14] In particular, we obtain the implied volatility for S&P 500 option contracts from the CBOE, calculated by inverting the Black-Scholes formula. We code days in terms of the percentage change in the implied volatility from the previous trading session.

[15] Notice that the proportion of undefined accounts is greater for outflows than for inflows. This is because outflows are relatively infrequent in our sample.

[16] Table 2 also indicates that the individual accounts classified as significant volatility chasers is higher (10.76%) than those classified as significant volatility avoiders (6.78%) although the proportion who display positive and negative volatility-chasing in general is about equal.

proportion measure is the dividing line between the two. [17] The results indicate that investor groups we identify typically display consistency over time. [18]

4.3 Aggregation

Once the investors have been grouped into different classes, we construct time series of their aggregate purchases and sales. In particular, in the case of the first set of groupings, investors are ranked in 50 groups in ascending order. Then, their purchases and sales are separately aggregated. This provides 50 time-series of both purchases and sales for each of the 5 groupings. Then, for each of the 50 categories we calculate the average of the absolute difference in percentage changes of purchases with respect to all the other 49 categories. That is, we do a pair-wise difference of each category with respect to all the others and take the average -i.e. a "High minus Low" for each category. We construct two time series, the first is the average of the time series for the first 25 categories and the second is the average of the time series of the last 25 categories separately considered. The resulting two time series' provide the first two factors. The other two factors are calculated analogously by using the sales. In an alternative specification we calculate the standard deviation of the value of these time series for the first 25 and the last 25 categories separately considered.

In the case of the grouping based on "consistent investors", we aggregate the purchases and sales of positive and negative feedback investors, defined on the basis of return and volatility. Each time series is composed of both purchases and sales of the investors belonging to the specific category. For example, the portfolio of negative return feedback investors ($NRFI$) has as its primitive four different time series: the purchases of the negative feedback investors identified on the basis of their sales, the purchases of the negative feedback investors identified on the basis of their purchases, the sales of the negative feedback investors identified on the basis of their sales and the purchases of the negative feedback investors identified on the

[17] We further restricted ourselves to accounts for which the probability level defined by the binomial test above exceeded 50% i.e. we only look at those who were more likely than not to be a feedback investor. Because of the infrequency of sales in the sample, there are relatively few feedback investors defined in terms of sales – not enough to perform the test.

[18] In particular, they show that daily return-feedback investors repeat both when they are defined in terms of purchases and when they are defined in terms of sales. In contrast, volatility-feedback investors do not seem consistent. Volatility positive feedback investors repeat only when defined in terms of purchases and not when defined in terms of sales. Volatility negative feedback investors repeat, but only when defined in terms of purchases, while the number of observation is not sufficient to draw any statistically significant conclusion when they are defined in terms of sales.

basis of their purchases.

Alternatively, we also construct a time series of the purchases and sales of feed-back investors separately considered, regardless of the direction of their reaction (positive or negative feedback investors). We therefore have purchases of return investors, purchases of volatility investors, sales of return investors, sales of volatility investors, net purchases (purchases minus sales) of return investors and net purchases of volatility investors. For example the time series of the purchases of return feed-back investors is made of the purchases of the negative return investors identified on the basis of their purchases, the purchases of the negative return investors identified on the basis of their sales, the purchases of the positive return investors identified on the basis of their purchases and the purchases of the positive return investors identified on the basis of their sales.

We then use the absolute differences in percentage changes in the transactions (either purchases, or sales or net purchases) of the different classes in order to construct the measure of dispersion of opinion.

One natural question is whether our measures of dispersion of opinion correlates with the standard measures of market uncertainty (implied volatility) and dispersion of opinion (open interest and trading volume). To test this, we regress our measure of dispersion of opinion on implied volatility, trading volume and open interest on the futures contracts written on the S&P500 index. The results show a strong correlation between our measures of dispersion of opinion and open interest and trading volume. No correlation, however, is found between it and implied volatility. This is consistent with early literature (Bessembinder, Chan and Seguin, 1996 and Christensen and Prabhala, 1998) that has identified open interest and implied volatility as proxies, respectively, for dispersion of beliefs and risk.

5 Empirical estimation

The next step is to test the explanatory power of our measures of dispersion of opinion. We adopt three approaches.

First, we consider the relationship between stock returns and dispersion of opinion using a standard asset pricing framework. In particular, we test the incremental explanatory power of the dispersion of opinion-related factors after having accounted for the explanatory power of the standard asset pricing factors.

A second approach relies on the fact that the model predicts that the relative explanatory power of the standard asset pricing factors and the dispersion of opinion depend on the overall market information asymmetry. An increase in overall information asymmetry raises the explanatory power of the weight of the dispersion of opinion-related factors and reduces the weight of the standard asset pricing factors. In particular, equation 6 suggests that the relation between stock returns and dispersion of opinion may be better characterized in terms of the existence of two different *regimes*. In the high-information-asymmetry-regime the role of the dispersion of opinion is high. In the low-information-asymmetry-regime the role of the dispersion of opinion is low. That is, the importance of the dispersion of opinion drops and standard asset pricing factors explain most of the stock returns. We can therefore explicitly condition on the two underlying regimes.

Finally, in the third approach we test whether measures of dispersion of opinion based on the behavior of "consistent investors" - i.e. investors who display consistent behavioral patterns - have explanatory power.

5.1 Dispersion of opinion and asset prices

Given that we want to assess the incremental explanatory power of the dispersion of opinion-related factors, we first orthogonalize these factors, by regressing them on the standard asset pricing factors. In particular, we estimate the auxiliary regression:

$$F_{D,t} = \theta + \delta F_{r,t} + \varepsilon_t. \tag{10}$$

$F_{r,t}$ are the standard asset pricing factors extracted by using stock returns and $F_{D,t}$ are the dispersion of opinion-related factors. Standard asset pricing factors are calculated through a standard factor extraction procedure applied on the 560 stock returns in the CRSP database that have been consecutively traded in the two-year period 1997-1998 with no missing observations.[19] The factor extraction is performed daily, by estimating loadings for each portfolio and portfolio weights via a principal component analysis performed on 90 days windows through the sample period.

Notice that, while standard asset pricing factors are autocorrelated over time, the dispersion of opinion is driven by learning errors which, by definition, should be independent over time and with average equal to zero. We therefore expect the factors based on past returns to be a good proxy of the standard asset pricing

[19]The reason we selected these stocks is that, given that we deal with investors into a S&P 500 index fund, we wanted to consider all the stocks that are part of the S&P 500 index or have analougous characteristics in terms of market capitalization.

factors and scarcely related to the dispersion of opinion-related ones. That is, this procedure should allow us to control for the relevant standard asset pricing factors driving the cross-section of returns.

Then, we test whether the dispersion of opinion-related factors have any additional incremental explanatory power to latent variables extracted from returns. This is done by estimating:

$$R_{i,t} = \alpha_i + \beta_i \varepsilon_t + \eta_{i,t}, \tag{11}$$

where ε_t are the residuals from the equation 10 and $R_{i,t}$ are the returns on portfolios of stocks.[20]

Equation 11 is estimated by using a standard Fama-MacBeth [FM] two-stage time-series cross-section test, applied to daily returns. We apply it to rolling intervals and daily updated betas. Given that we need a 90-day rolling window to estimate the factors, our sample consists of 412 observations (March 1997-December 1998). This generates sets of betas that are then used as explanatory variables in the second step of the procedure. In order to overcome the potential problems of lead-lag effects due to asynchronous trading with daily data, we apply the Dimson-Marsh correction using two days of leads and lags. In stage 2, we regress portfolio returns on betas each day following the estimation period and save the resulting adjusted *Adjusted RSquare*. [21]

If our hypothesis is correct, we expect βs to have additional explanatory power. This additional power, based on an information set orthogonal to the one contained in past stock returns, allows us the gauge the role played by dispersion of opinion on asset prices.

The results, reported in Tables 3, appear to support our hypothesis, displaying a significant additional explanatory power of the dispersion of opinion-related factors. This holds for all the specifications considered. The values of the *Adjusted RSquare* are very similar, regardless of the criterion used to classify investors. This also provides a good robustness check that the results are not biased by the implicit endogeneity of the criterion employed to group the investors. Indeed, the results do not differ even if some criteria are more subject than others to the endogeneity bias.

[20] In particular out of the selected sample we create 20 portfolios, each containing 28 stocks, ranked by market capitalization.

[21] For a robustness check, we repeat the same experiment for several different time horizons: that is the next 5, 10, 15, 20 and 40 days following the estimation period. Longer horizon returns show similar results. Given that all the results agree, we report only the standard one based on time t.

It is also worth noting that the explanatory power of the dispersion of opinion-related factors is not dissimilar from the one of the standard asset pricing factors (on average 9.5-10.5% as opposed to the 12.6%).

As a robustness check, we also estimate:

$$R_{i,t} = \alpha_i + \beta_i \varepsilon_t + \gamma_i F_{r,t} + \eta_{i,t}. \tag{12}$$

If the dispersion of opinion-related factors load on a different source of uncertainty that the standard asset pricing factors, we expect them to increase the explanatory power in equation 12. Also we can explicitly test for the significance of the incremental explanatory power by comparing the *Adjusted RSquare* of the specification with both sets of factors to the specification with only the standard asset pricing factors. A t-test for distribution with unequal variances is therefore performed. The results are reported in Table 4. They support the hypothesis that the dispersion of opinion-related factors provide a significant improvement in the explanatory power.

5.2 Dispersion of opinion and regimes

A second way of assessing the role played by the dispersion of opinion exploits the fact that the impact of the dispersion of opinion differs depending on the degree of overall information asymmetry. We therefore explicitly condition on the type of underlying information regime. In order to identify the two regimes of informational asymmetry, we use a standard Markov-switching technique (Hamilton 1990). We assume the existence of an unobserved random variable (s_t) that takes the values *1* or *2* according to which regime the process is in at time t: one characterized by high overall information asymmetry in the market (H) and one characterized by low overall information asymmetry in the market (L). The probability law governing the shifts between high and low states is represented by a two-state Markov chain such that:

$$\begin{aligned} P(s_t &= H|s_{t-1} = H) = P_{11}, \ P(s_t = L|s_{t-1} = H) = 1 - P_{11}, \\ P(s_t &= H|s_{t-1} = L) = 1 - P_{22} \text{ and } P(s_t = L|s_{t-1} = L) = P_{22}. \end{aligned}$$

We may therefore rewrite equation 6 as:

$$R_{H,t} = \alpha_H + \beta_H F_t + \gamma_H Disp_t + \varepsilon_{H,t} \text{ and } R_{L,t} = \alpha_L + \beta_L F_t + \gamma_L Disp_t + \varepsilon_{L,t}, \tag{13}$$

where $R_{H,t}$ and $R_{L,t}$ are the returns at period t, conditional on the two regimes (high and low uncertainty respectively) on the S&P500 index. $Disp_t$ proxies for the dispersion of opinion-related factors. We consider two sets of dispersion of opinion related-factors: the ones based on purchases and the ones based on sales. They have been constructed as described in the previous section. F_t proxies for the standard asset pricing factors, that is all the components of returns not related to dispersion of opinion. They are defined using some standard information variables such as dividend yield, yield on long term corporate bonds (AAA quality), yield on junk bonds, yield on the Treasury Bills (Ferson and Harvey, 1999).

We also consider specifications including the returns on the S&P 500 lagged one period and the market overall trading volume. These variables have been selected in order to provide the largest possible set available to the investors in the market and which can be used to infer the standard asset pricing factors. In the case overall trading volume is included among the information variables, F_t is orthogonalized by regressing it on volume and taking the residuals. This should provide a better measure of the dispersion of the dispersion of opinion not captured by the overall trade.

Also, given that (see Theorem 2) both informed and uninformed investors' trading are affected by the fundamentals (D_t), overall volume should provide an indirect proxy for it. This suggests a role for trading volume, different from the standard identification with dispersion of opinion. The hypotheses underlying the statistical model are standard: the error term in the observation equation, ε_t, is assumed to be i.i.d. normal. An algorithm based on the EM principle is applied. Given an ML-estimate of the vector of the parameters the optimal inference on the hidden Markov process is found by iteration.

The results are reported in Table 5. All the specifications (with and without volume, with and without lagged returns) agree. The value and the statistical significance of the dispersion of opinion-related factors depend on the type of regime. In one regime (II Regime), the γs on the dispersion of opinion-related factors are significant and positive, both in the case of purchases and sales. [22] In the other regime (I Regime), on the contrary, the average significance disappears. While the measures of dispersion of opinion based on purchases are on average scarcely significant and positive, the ones based on sales are significant and negative. Given that their effects tend to offset each other, the aggregate impact of the dispersion of opinion-related factors differs in the two regimes: strongly positive in the second

[22] When the additional controls of lagged past returns and overall trading volume are added, the significance drops and in a few cases disappears. On average, however, they are strongly significant and positive.

regime, on average null or slightly negative in the first one. This provides evidence that dispersion of opinion-related factors may affect stock returns in a different way, depending on the degree of informational asymmetry in the market.

5.3 Dispersion of opinion and consistent investors

The previous results suggest that our measure of dispersion of opinion has incremental explanatory power with respect to standard asset pricing factors. We now focus on specific groups of investors and try to relate dispersion of opinion to their behavior. In particular, we use the purchases and sales of a subset of investors, (the "consistent investors" defined above) as a measure of dispersion of opinion. We start from equation:

$$R_{i,t} = \alpha_i + \beta_i F_{r,t} + \gamma F_{D,t} + \varepsilon_{i,t}, \tag{14}$$

where $F_{r,t}$ are the standard asset pricing factors and $F_{D,t}$ are the dispersion of opinion-related factors. We consider eight standard asset pricing factors (four factors are extracted from past returns and four are based on the investors' flows orthogonalized by regressing them on the first four factors) and four dispersion of opinion-related factors.

We focus on two specifications. In the first one the measure of dispersion of opinion is constructed by using both the purchases and sales for positive and negative feedback investors. That is, the flows of positive and negative feedback investors are separately considered investors are separately. In the second specification, the measure of dispersion of opinion is based on purchases and sales separately considered for a generic feedback investor. That is, factors are constructed by considering purchases and sales separately, but investors are aggregated regardless of the "direction" of the reaction (positive or negative).

We report the adjusted r_t^2 as well as p-values of the test whether the means of the r_t^2 of the regressions with the dispersion of opinion-related factors are statistically different from the means of the adjusted r_t^2 estimated using only the standard asset pricing factors.

Unlike equation 11 in the previous section, here we run a direct horse-race between standard asset pricing factors and dispersion of opinion. We also impose a more demanding condition by controlling for the residual explanatory power of investor flows, after they have been orthogonalized with respect to past returns. As before, our goal is to test whether adding our measures of dispersion of opinion increases the explanatory power.

The results, reported in Table 6, show a strong and significant increase in the explanatory power of the regression due to the addition of the factors based on dispersion of opinion. This holds for all the specifications that have been considered.

In a second approach, we add a third stage to the FM procedure. We calculate the time-series of the residuals from the daily cross-sectional FM regressions on factors constructed using returns and flows but not dispersion of opinion as in equation 14.[23] Then, we regress these residuals on our measure of opinion dispersion as:

$$\varepsilon_{i,t}, = \alpha + \beta F_{D,t} + \eta_t, \tag{15}$$

where different specifications of the dispersion of opinion-related factors ($F_{D,t}$). [24] This allows us to see whether the dispersion of opinion explains the residuals. Also, this specification lets us see the sign of the relationship between dispersion of opinion-related factors and residuals. Given that the dispersion of opinion-related factors tend to have higher explanatory power at the time when the standard asset pricing factors have a relatively lower power, we would expect: $\beta < 0$.

The results, reported in Table 7, show a significant negative relationship between explanatory power in the FM regression and measures of dispersion of opinion. This suggests that not only does dispersion of opinion significantly increase the explanatory power of the FM regressions, but also it does this exactly at the times where the standard factors provide a worse fit.

[23]The 8 factors are: four extracted from past returns (standard market factors), four based on the investors' flows orthogonalized by regressing them on the first four factors (behavioral factors)

[24]We consider alternative specifications that differ depending on the type of flows we use to construct these four actors. We consider either the purchases and sales of the investors identified in terms of positive (positive return or volatility investors) and negative (negative return or volatility investors) reactions, or the purchases or sales separately considered of the investors defined on in terms of the event they react to (return and volatility investors). Also the case when only the first four factors extracted from past returns is considered ("Return"). The dispersion of beliefs-related factors are constructed using the same way of aggregating the transactions (purchases and sales) of different classes of rational investors used to build the previously defined fundamentals-related factors. They are constructed as the absolute differences between percentage changes of positive and negative feedback investor flows, both defined in terms of return and volatility. The factors are constructed using the flows (both purchases and sales) of positive and negative feedback investors, defined on the basis of return and volatility. Each single portfolio is composed of the percentage changes in both purchases and sales of the investors belonging to the specific category. For example, the portfolio of negative return feedback investors (NRFI) is made of four components: a measure of dispersion constructed by using the purchases of the negative feedback investors identified on the basis of their sales, a measure of dispersion constructed by using the purchases of the negative feedback investors identified on the basis of their purchases, a measure of dispersion constructed by using the sales of the negative feedback investors identified on the basis of their sales and a measure of dispersion constructed by using the purchases of the negative feedback investors identified on the basis of their purchases.

6 Conclusion

In this paper, we study the way heterogeneity of trade among investors affects stock returns.

Focusing on the pricing of the aggregate U.S. equity market, we test its empirical implications by using panel data of index funds' investors. We show that dispersion of opinion, proxied by the heterogeneity of trade among investors, may explain part of the returns not accounted for by the standard asset pricing factors. In particular, we show that the explanatory power of the dispersion of opinion appears to increase at the very time pricing models based on standard asset pricing factors fare worse.

These results suggest a way of addressing the deficiencies of the standard pricing models that require us to account for "extraneous risk", that is sources of risks not directly traceable to the standard asset pricing factors.

They also stress the potential value of the information contained in investors' flows, mostly orthogonal to the information contained in returns, and suggest a way of exploiting it.

References

[1] Balduzzi P., G. Bertola, S. Foresi,1995, " Asset Price Dynamics and Infrequent Feedback Trades" Journal of Finance, Vol. 50, No. 5. (Dec.), pp. 1747-1766.

[2] Barberis N., A. Shleifer and R.Visny, 1998, " A model of Investor Sentiment," Working Paper, Havard Economics Department, 1998.

[3] Bessembinder, H., Chan, K. and P.J. Seguin, 1996, "An empirical examination of information, differences of opinion, and trading activity", *Journal of Financial Economics*, 40, 105-134.

[4] Biais B. and P. Bossaerts, 1999, " Asset prices and trading volume in a beauty contest", The Review of Economic Studies, 65, 307-340.

[5] Bossaerts P., Plott C. and Zame W., 2000, "Prices And Allocations In Financial Markets: Theory and Evidence" Caltech W.P.

[6] Brennan M. J., Jegadeesh N and Swaminathan B., 1993, "Investment analysis and the adjustment of stock prices to common information", The Review of Financial Studies, 799-824.

[7] Brennan M. J., Chordia T. and Subrahmanyam., 1998, "Alternative factor specifications, security characteristics, and the cross section of expected returns", The Journal of Financial Economics, 49, 345-373.

[8] Campbell J.Y, and Kyle A.S., 1993, "Smart money, noise trading and stock price behavior", Review of Economic Studies, 60, 1-34.

[9] Campbell J.Y., Grossman S. and J. Wang., 1992 "Trading volume and serial correlations in stock returns", NBER Working Paper #4193

[10] Chan L. and J. Lakonishok, 1993, "Institutional Trades and Intraday Stock Price Behavior", Journal of Financial Economics, 33(2), 173-99.

[11] Chan L., Jegadeesh N. and J. Lakonishok, 1996, "Momentum strategies", Journal of Finance, 51(5), 1681-1711.

[12] Daniel K., D. Hirshleifer and A. Subrahmanyam, 1998, "A theory of Over-confidence, self-Attribution and Security Market Over and Under reactions," working paper, UCLA.

[13] De Bondt W. F. M. and R. Thaler, 1985, "Does the Stock Market Overreact?" Journal of Finance, Vol. 40, No. 3, (Jul), pp. 793-805.

[14] De Long J. B., A. Shleifer, L. H. Summers and R. J. Waldmann, 1990, "Noise trader risk in financial markets", Journal of Political Economy, 98(4), 703-738.

[15] Detemple J. and S. Murthy, 1994, "Intertemporal asset pricing with heterogeneous opinion", Journal of Economic Theory, 62, 294-320.

[16] Ferson, W. E., (1990), Are the Latent Variables in Time-Varying Expected Returns Compensation for Consumption Risk? Journal of Finance.

[17] Ferson, W. E. and Harvey, C. R (1999) Conditioning Variables and the Cross Section of Stock Returns Journal of Finance;

[18] Ferson, W. E. and Harvey C.R., (1995) Predictability and Time-Varying Risk in World Equity Markets, Chen,-Andrew-H., ed. Research in finance, JAI Press.

[19] Fung, W. and D. Hsieh, 1998, "Empirical characteristics of dynamic trading strategies: The case of hedge funds," Review of Financial Studies 10: (2) 275-302 .

[20] Goetzmann W. and M. Massa (1999) "Index funds and stock market growth", Mimeo, Yale University.

[21] Gompers P. and A. Metrick, 1998, "How are Large Institutions Different From Other Investors? Why do These Differences Matter?" NBER Working Paper.

[22] Grinblatt M., Titman, S. and R. Wermers, 1995, " Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior", American-Economic-Review; 85(5), 1088-1105.

[23] Grinblatt M. and M. Keloharju, 1999, "What makes investors trade?", Mimeo UCLA

[24] Grossman S., 1995, "Dynamic asset allocation and the informational efficiency of the markets", Journal of Finance, 50, 773-787.

[25] Harris, M. and A.Raviv, (1993) "Differences of opinion make a horse race", The Review of Financial Studies, 6, 473 - 506

[26] Hau H., (1998), "Competitive entry and endogenous risk in the foreign exchange market", The Review of Financial Studies", 11(4), 757-787

[27] He, H. and Wang, J., 1995, "Differential Information and Dynamic Behavior of Stock Trading Volume", Review-of-Financial-Studies,8(4), 919-72.

[28] Hirshleifer D., A. Subrahmanyam, S. Titman, 1994, "Security Analysis and Trading Patterns when Some Investors Receive Information Before Others," Journal of Finance, Vol. 49, No. 5. (Dec), pp. 1665-1698.

[29] Hong H. and J. Stein, 1999, "A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets," Forthcoming, Journal of Finance.

[30] Jeegadesh N. and S. Titman, 1995, "Overreaction, delayed reaction and contrarian profits", The Review of Financial Studies, 8(4), 973-993.

[31] Kandel, E. and N. Pearson, 1995, "Interpretation of public signals and trade in speculative markets", *Journal of Political Economy*, 103, 831-872.

[32] Keim D. and A. Madhavan, 1995, "Anatomy of the trading process. Empirical evidence on the behavior of institutional traders," Journal of Financial Economics 37(3) p. 371-398.

[33] Kraus A. and M. Smith, 1989, "Market created risk", The Journal of Finance, 44, 557-569.

[34] Lakonishok J. and A. Shleifer, 1992, "The Impact of Institutional Trading on Stock Prices", Journal of Financial Economics, 32(l), 23-43.

[35] Lakonishok J., A. Shleifer and R. W. Vishny, 1991, "Do Institutional Investors Destabilize Stock Prices? Evidence on Herding and Feedback Trading", NBER Working Paper: 3846.

[36] Lo A.W. and A.C. MacKinlay, 1990, "When are Contrarian Profits Due to Stock Market Overreaction?" Review of Financial Studies, Vol. 3, No. 2. , pp. 175-205.

[37] Lo, A and Wang, 2000, "Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory", Review-of-Financial-Studies; 13(2), 257-300.

[38] Naik N., 1997, "On Aggregation of Information in Competitive Markets: The Dynamic Case" Journal of Economic Dynamics and Control, 21(7), 1199-1227.

[39] Shalen C. T., 1993, "Volume, volatility and the dispersion of opinion", Review of Financial Studies, 6, 405-434.

[40] Sias R.W. and L.T. Starks, (1996) "Return Autocorrelation and Institutional Investors", Journal of Financial Economics, 46, 103-31.

[41] Wang, J., 1993, "A Model of Intertemporal Asset Prices under Asymmetric Information", Review-of-Economic-Studies, 60(2), 249-82..

[42] Williams J., 1977, " Capital asset prices with heterogeneous opinion", Journal of Financial Economics, 219-277.

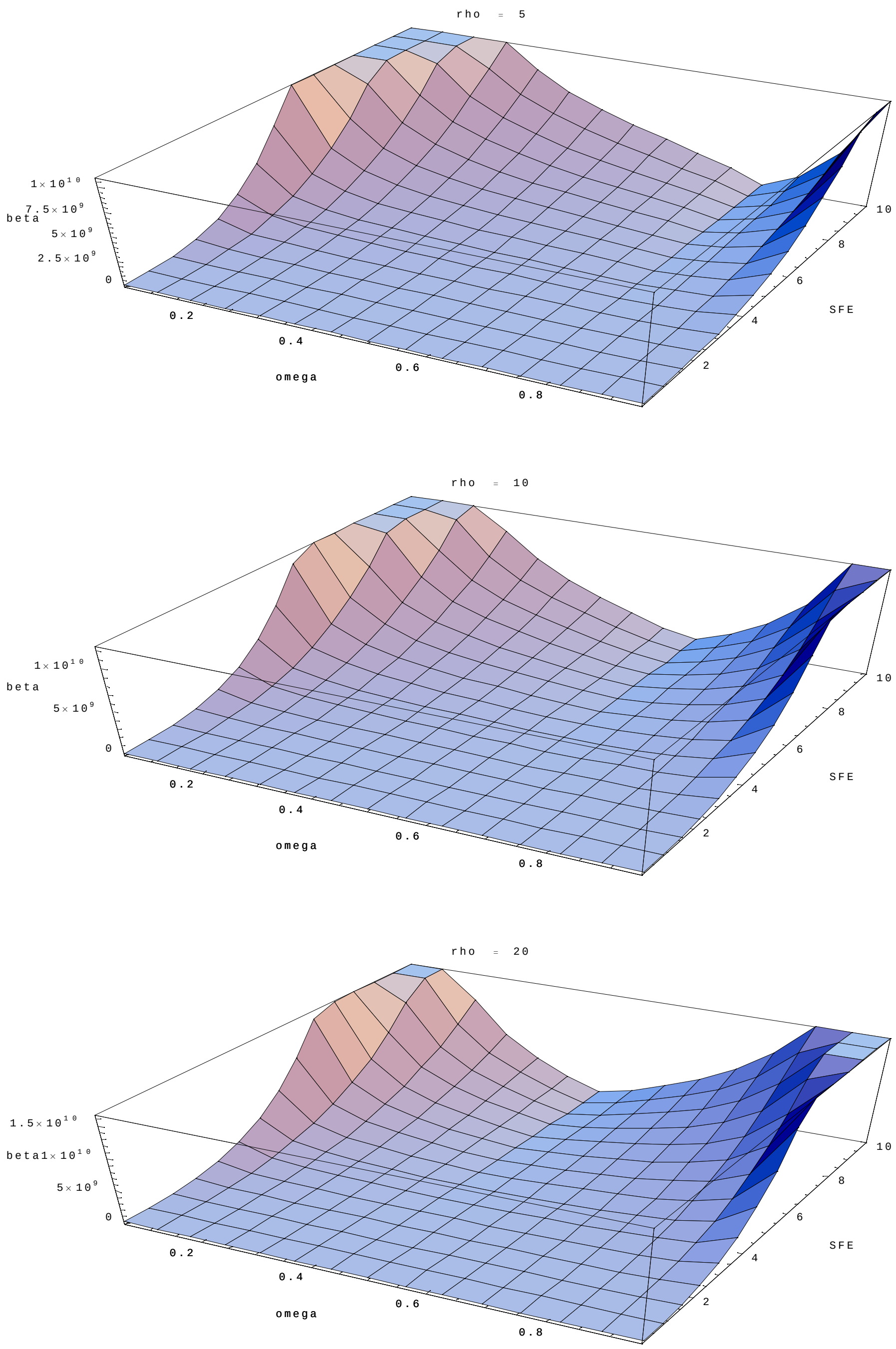

Figure 1: **Impact of dispersion of beliefs on stock returns (β).** The coefficient on the dispersion of beliefs is: $\beta = \frac{(1-\omega e_0)}{\omega e_0}$, where $r = 0.05, a_\Pi = 0.2$, $\sigma_D = 0.046$ and $\sigma_\Pi = 0.018$.

Table 1: Descriptive Statistics

Investors are grouped into 4 categories, on the basis of some institutional differences. The Individuals Accounts include: Administrator, Individual, Non-Prototype Individual, Sole Proprietorship, and Personal Representative. The Tax-benefited Accounts include Traditional IRA, UTMA, Rollover IRA, Sep-IRA, Joint-WROs, Money Purchase Keogh, Non-Prototype IRA, ROTH IRA, Simple IRA and PS Voluntary Keogh. The Fiduciary and Trusts Accounts include the Conservator, Executor, Fiduciary, Guardian, Transfer on Death-Individual, Trust: under Agreement, Trust under Indenture, Trust under Will. The Groups Accounts include the Bank, Religious Organisation, Joint CP, Corporation, Investment Club, Professional Corp., Partnership, Joint TIC, Joint TBE, Unincorporated Association, UGMA, Professional Association. Running Balance is constructed as the average holdings standardised by the amount of time they are held. Turnover is calculated as the absolute sum of purchases and sales (expressed in terms of number of shares) in the fund divided by the average running balance. Investor Profit Ratio is calculated as the ratio between the terminal value of the sum of the inflows and outflows each accrued at the return on the index fund and the terminal value of a buy and hold strategy.

		Individuals	Tax-benefited Accounts	Fiduciary and Trust	Groups	Total
Number of Accounts		16,185	66,903	5,493	2,179	90,768
Number of Transactions		51,864	185,059	15,558	7,119	259,614
Percentage of Purchases		0.82	0.82	0.83	0.81	0.82
Percentage of Sales		0.18	0.17	0.16	0.18	0.17
Running Balance	Mean	400	254	584	1341	327
(in number of shares)	Median	170	116	244	195	134
	S.Dev	1,106	665	2,061	2,259	3,617
Turnover Ratio	Mean	30.18	16.55	202.35	15.68	30.23
	Median	1.12	1.02	1.03	1.11	1.03
	S.Dev	997.19	608.69	12,556.5	1,250	3,160
Investor Profit Ratio	Mean	1.17	1.19	1.97	1.21	1.23
	Median	0.85	0.88	0.85	0.84	0.87
	S.Dev	6.14	13.14	68.41	5.05	20.44
Number of Transactions	Mean	3.20	2.76	2.83	3.26	2.86
	Median	2.00	2.00	2.00	2.00	2.00
	S.Dev	4.30	3.82	4.79	5.09	4.01

Table 2: Investor Typology

Negative return feedback investors (NRFI) are defined as the investors who invest in the fund when the daily return of the index of the previous day is negative and positive return feedback investors (PRFI) are defined as the investors who invest in the fund when the daily return of the index of the previous day is positive. Negative volatility feedback investors (NVFI) are defined as the investors who invest in the fund when the volatility of the day before the investment is decreasing with respect to the previous day positive volatility feedback investors (PVFI) are defined as the investors who invest in the fund when the volatility of the day before the investment is increasing with respect to the previous day. Volatility is the implied volatility on the option on the SP500 as defined using the Black-Sholes pricing formula. Consistent agents are identified on the basis of their *systematic* behavior. A small sample test of equality between the distribution of investors' behavior and market returns based on the binomial distribution is applied and the investors with a statistic greater than 10% have been identified as consistent investors. All the cases where the test is equal to zero or is not defined are called "undefined". Only accounts with at least 3 transactions are considered.

All Accounts

		Individuals (%)		Tax-benefited Accounts (%)		Fiduciary and Trust (%)		Groups (%)		Total (%)	
		Purch.	Sales	Purch.	Sales	Purch.	Sales	Purch.	Sales	Purch.	Sales
PRFI	α>0.1	1.19	0.10	1.04	0.12	1.06	0.11	1.51	0.14	1.08	0.11
	0.5>α>0.1	11.16	2.26	9.97	1.77	9.72	1.80	13.35	2.34	10.25	1.87
Undef.		63.89	85.32	69.46	87.11	66.76	88.26	61.22	85.87	68.10	86.83
NRFI	0.5>α>0.1	20.87	11.99	17.38	10.76	19.55	9.58	20.70	11.11	18.22	10.91
	α>0.1	2.90	0.34	2.15	0.24	2.91	0.25	3.21	0.55	2.36	0.27
VPFI	α>0.1	1.84	0.14	1.28	0.08	1.20	0.15	1.42	0.18	1.38	0.09
	0.5>α>0.1	12.65	2.37	10.47	1.72	12.12	1.67	13.45	1.97	11.03	1.84
Undef.		66.76	87.00	71.46	88.42	68.60	89.50	65.08	86.97	70.29	88.20
NPFI	0.5>α>0.1	17.71	10.41	15.94	9.70	17.29	8.54	19.00	10.83	16.41	9.79
	α>0.1	1.05	0.08	0.85	0.08	0.78	0.15	1.06	0.05	0.88	0.08

Accounts with more than 8 transactions

		Purch.	Sales	Purch.	Sales	Purch.	Sales	Purch.	Sales	Purch.	Sales
PRFI	α>0.1	10.40	0.89	12.38	1.26	11.07	2.29	17.16	1.49	11.99	1.24
	0.5>α>0.1	35.38	9.33	38.92	7.96	40.46	9.92	39.55	10.45	38.21	8.44
Undef.		11.11	74.40	11.94	73.51	12.98	74.81	7.46	73.13	11.68	73.77
NRFI	0.5>α>0.1	27.11	13.78	23.85	15.28	21.37	11.07	21.64	11.19	24.40	14.60
	α>0.1	16.00	1.60	12.91	1.99	14.12	1.91	14.18	3.73	13.71	1.94
VPFI	α>0.1	16.44	1.42	14.93	0.88	14.89	1.91	12.69	2.99	15.21	1.11
	0.5>α>0.1	28.98	8.62	30.55	8.87	30.15	8.02	29.10	8.21	30.13	8.75
Undef.		15.56	76.62	12.91	75.83	15.65	77.48	11.19	79.10	13.61	76.18
NPFI	0.5>α>0.1	28.62	12.36	30.49	13.37	29.77	9.54	37.31	8.96	30.21	12.82
	α>0.1	10.40	0.98	11.12	1.05	9.54	3.05	9.70	0.75	10.83	1.13

Table 3: Stock returns and dispersion of opinions
(incremental explanatory power of dispersion of opinions)

The table reports the means of the *Adjusted R^2* from the second stage of a Fama-MacBeth procedure based *only* on 4 dispersion of opinions-related factors. The *dispersion of opinions-related factors* are constructed by identifying the transactions (purchases and sales) of different classes of investors. Investors are identified in terms of the amount of money invested in the index fund on average (Average Holdings), the money the have invested at the end of the period (Running Balance), the dispersion of the holdings over time (Holding Dispersion), their frequency of trading (Number of Transactions and Turnover). Average Holdings are defined as the number of shares the investor has in the fund multiplied by the length of time they are held, the Dispersion of Holdings is the standard deviation of the holdings over time. Turnover is calculated as the absolute sum of purchases and sales in the fund divided by the average running balance and Running Balance is constructed as the average holdings standardized by the amount of time they are held. Investors are then ranked in 50 groups in ascending order and their purchases and sales are separately aggregated. This provides 50 time-series of both purchases and sales for each of the 6 groupings. Then for each of the 50 categories we calculate the absolute difference in percentage changes of purchases with respects to all the other 49 categories. We calculate the average value of these time series for the first 25 and the last 25 categories separately considered (*Specification I*). The resulting time series provide the first two factors. The other two factors are calculated analogously by using the sales. In an alternative specification we calculate the standard deviation of the value of these time series for the first 25 and the last 25 categories separately considered (*Specification II*). *The dispersion of opinions-related factors are then orthogonalized* by regressing them on the standard asset pricing factors. The *standard asset pricing factors* are extracted from past returns. In particular, we consider the regularly traded individual securities in the U.S. market. Loadings for each portfolio and portfolio weights are estimated via a principal component analysis performed on over-lapping 90 days windows through the sample period. The factors are extracted and loadings estimated using leading rolling windows. For the returns, we take the 560 stocks in the CRSP database that have been consecutively traded in the two-year period 1997-1998 with no missing observations. We then create 20 portfolios each containing 28 stocks, ranked by market capitalization. A Dimson-Marsh correction using two days of leads and lags is applied to control for potential lead-lag effects due to asynchronous trading. The factor extraction and the estimation of the betas are updated each day in the sample, following the initial 90-day estimation period. Thus, betas are allowed to vary through time. In stage 2, we regress portfolio returns on betas for each day following the estimation period. For each day a cross-section over the 20 portfolios is estimated. We report the mean values of the *Adjusted R^2* of such a cross-section, averaged over time.

	I Specification	**II Specification**
Full Set of Factors. Classification based on:		
Holding Dispersion	0.103	0.093
Running Balance	0.102	0.094
Average Holdings	0.100	0.094
Number of Transactions	0.106	0.094
Portfolio Turnover	0.105	0.094

Table 4: Stock returns and dispersion of opinions
(statistical significance of incremental explanatory power)

The table reports the means of the R^2 from of the second stage of a Fama-MacBeth procedure with 8 factors: four standard asset pricing factors and four dispersion of opinions-related factors. *The standard asset pricing factors* are extracted from past returns. In particular, we consider the regularly traded individual securities in the U.S. market. Loadings for each portfolio and portfolio weights are estimated via a principal component analysis performed on over-lapping 90 days windows through the sample period. The factors are extracted and loadings estimated using leading rolling windows. For the returns, we take the 560 stocks in the CRSP database that have been consecutively traded in the two-year period 1997-1998 with no missing observations. We then create 20 portfolios each containing 28 stocks, ranked by market capitalization. The *dispersion of opinions-related factors* are constructed by identifying the transactions (purchases and sales) of different classes of investors. The classes are determined by grouping the accounts on the basis of the characteristics of the investors. Investors are identified in terms of the amount of money invested in the index fund on average (Average Holdings), the money the have invested at the end of the period (Running Balance), the dispersion of the holdings over time (Holding Dispersion), their frequency of trading (Number of Transactions and Turnover). Average Holdings are defined as the number of shares the investor has in the fund multiplied by the length of time they are held, the Dispersion of Holdings is the standard deviation of the holdings over time. Turnover is calculated as the absolute sum of purchases and sales in the fund divided by the average running balance and Running Balance is constructed as the average holdings standardized by the amount of time they are held. Investors are then ranked in 50 groups in ascending order and their purchases and sales are separately aggregated. This provides 50 time-series of both purchases and sales for each of the 6 groupings. Then for each of the 50 categories we calculate the absolute difference in percentage changes of purchases with respects to all the other 49 categories. We calculate the average value of these time series for the first 25 and the last 25 categories separately considered (*Specification I*). The resulting time series provide the first two factors. The other two factors are calculated analogously by using the sales. In an alternative specification we calculate the standard deviation of the value of these time series for the first 25 and the last 25 categories separately considered (*Specification II*). The dispersion of opinions-related factors are then orthogonalized by regressing them on the first four factors (standard asset pricing factors). A Dimson-Marsh correction using two days of leads and lags is applied to control for potential lead-lag effects due to asynchronous trading. The factor extraction and the estimation of the betas are updated each day in the sample, following the initial 90-day estimation period. Thus, betas are allowed to vary through time. In stage 2, we regress portfolio returns on betas for each day following the estimation period. For each day a cross-section over the 20 portfolios is estimated. We report the mean values of the *Adjusted* R^2 of such a cross-section, averaged over time. We consider the case based only on the standard asset pricing factors (4 factors) and the case based on both standard asset pricing factors and dispersion of opinions factors (Full set of 8 factors). For the specification inclusive of all the 8 factors we also report the *P-Value* of the t-test testing whether the means of the *Adjusted* R^2 estimated using full set of factors are statistically different from the ones estimated using only the standard asset pricing factors.

	I Specification		**II Specification**	
	Mean	**P Value**	**Mean**	**P Value**
Standard asset pricing factors only	0.126	-	0.126	-
Full Set of Factors. Classification based on:				
Holding Dispersion	0.172	0.0001	0.173	0.0001
Running Balance	0.160	0.0010	0.154	0.0080
Average Holdings	0.163	0.0001	0.172	0.0001
Number of Transactions	0.174	0.0001	0.168	0.0001
Portfolio Turnover	0.176	0.0001	0.160	0.0020

Table 5: Impact of Dispersion of opinions in different regimes

A Markov-switching regime model is estimated for the specification $R_t = \alpha + ß_1 DB_{1,t} + ß_2 DB_{2,t} + \gamma LTY_t + \delta JY_t + \zeta STY_t + \mu R_{t-1} + \nu DY_t + \theta V_t + \varepsilon_t$, where R_t is the daily return on the S&P500 index, LTY_t is the yield on long term corporate bond, JY_t is the yield on Junk bond, STY_t is the yield on the T-Bills, DY_t is the dividend yield of the Index and Vt is the overall trading volume on the S&P 500. DB_t is the measure of dispersion of opinions. It is constructed by identifying the transactions (purchases and sales) of different classes of investors. The classes are determined by grouping the accounts on the basis of the characteristics of the investors. Investors are identified in terms of the amount of money invested in the index fund on average (Average Holdings), the money the have invested at the end of the period (Running Balance), the dispersion of the holdings over time (Holding Dispersion), their frequency of trading (Number of Transactions and Turnover). Average Holdings are defined as the number of shares the investor has in the fund multiplied by the length of time they are held, the Dispersion of Holdings is the standard deviation of the holdings over time. Turnover is calculated as the absolute sum of purchases and sales in the fund divided by the average running balance and Running Balance is constructed as the average holdings standardized by the amount of time they are held. Investors are then ranked in 50 groups in ascending order and their purchases and sales are separately aggregated. This provides 50 time-series of both purchases and sales for each of the 6 groupings. Then for each of the 50 categories we calculate the absolute difference in percentage changes of purchases with respects to all the other 49 categories. We calculate the average value of these time series for 50 categories. The resulting time series provides the first factor ($DB_{1,t}$). The other factor is calculated analogously by using the sales ($DB_{2,t}$). The dispersion of opinions-related factors are then orthogonalized by regressing them on the first four factors (standard asset pricing factors). We consider alternative specifications that differ, depending on whether lagged returns or contemporaneous trading volume on the S&P500 is included. In the latter case, the measure of dispersion of opinions is previously orthogonalized by regressing it on the trading volume itself. The flows (purchases and sales) have been standardized by dividing them by 100,000.

I Specification

($R_t = \alpha + ß_1 DB_{1,t} + ß_2 DB_{2,t} + \gamma LTY_t + \delta JY_t + \zeta STY_t + \nu DY_t + \varepsilon_t$,)

	ß$_1$				**ß$_2$**			
	I Regime		**II Regime**		**I Regime**		**II Regime**	
	Value	**Tstat**	**Value**	**TStat**	**Value**	**TStat**	**Value**	**TStat**
Full Set of Factors.								
Classification based on :								
Holding Dispersion	0.02	0.60	0.30	2.24	-0.61	-7.15	0.67	2.60
Running Balance	0.03	0.88	0.91	3.91	-0.55	-6.98	0.91	2.80
Average Holdings	0.03	0.83	0.90	4.09	-0.60	-6.80	0.85	2.45
Number of Transactions	0.02	0.61	0.83	4.56	-0.56	-7.03	0.84	2.65
Portfolio Turnover	0.02	0.73	0.84	4.04	-0.57	-6.91	0.91	2.96

II Specification

$(R_t = \alpha + ß_1\ DB_{1,t} + ß_2\ DB_{2,t} + \gamma LTY_t + \delta JY_t + \zeta STY_t + \nu\ DY_t + \theta V_t + \varepsilon_t,)$

	ß1				ß2			
	I Regime		**II Regime**		**I Regime**		**II Regime**	
	Value	**Tstat**	**Value**	**TStat**	**Value**	**TStat**	**Value**	**TStat**
Full Set of Factors.								
Classification based on :								
Holding Dispersion	0.01	1.30	0.01	0.73	-0.05	-5.34	0.07	2.82
Running Balance	0.01	1.53	0.02	0.69	-0.05	-5.52	0.06	2.21
Average Holdings	0.01	1.83	0.14	2.00	-0.02	-2.09	0.10	6.41
Number of Transactions	0.00	1.01	0.03	2.26	-0.05	-6.27	0.06	2.10
Portfolio Turnover	0.01	1.41	0.01	0.75	-0.05	-5.36	0.07	2.68

III Specification

$(R_t = \alpha + ß_1\ DB_{1,t} + ß_2\ DB_{2,t} + \gamma LTY_t + \delta JY_t + \zeta STY_t + \mu R_{t-1} + \nu\ DY_t + \varepsilon_t,)$

	Flows-based Dispersion of Opinions							
	ß1				ß2			
	I Regime		**II Regime**		**I Regime**		**II Regime**	
	Value	**TStat**	**Value**	**Tstat**	**Value**	**TStat**	**Value**	**TStat**
Full Set of Factors.								
Classification based on :								
Holding Dispersion	0.02	0.55	0.21	1.70	-0.63	-6.69	0.94	3.26
Running Balance	0.03	0.91	0.81	4.15	-0.59	-7.15	0.73	2.36
Average Holdings	0.03	0.78	0.42	1.82	-0.68	-7.40	0.53	1.78
Number of Transactions	0.02	0.66	0.30	2.49	-0.63	-7.65	0.66	2.43
Portfolio Turnover	0.02	0.75	0.34	2.31	-0.63	-7.59	0.71	2.58

Table 6: Dispersion of opinions and consistent investors
(incremental explanatory power)

The table reports the means of the R^2 from the second stage of a Fama-MacBeth procedure with 12 factors: eight standard asset pricing factors (four extracted from past returns, four based on the investors' flows orthogonalized by regressing them on the first four factors) and four dispersion of opinions-related factors. The standard asset pricing factors are extracted from past returns. In particular, we consider the regularly traded individual securities in the U.S. market. Loadings for each portfolio and portfolio weights are estimated via a principal component analysis performed on over-lapping 90 days windows through the sample period. The factors are extracted and loadings estimated using leading rolling windows. For the returns, we take the 560 stocks in the CRSP database that have been consecutively traded in the two-year period 1997-1998 with no missing observations. We then create 20 portfolios each containing 28 stocks, ranked by market capitalization. The dispersion of opinions-related factors are constructed by identifying the transactions (purchases and sales) of different classes of consistent investors. They are constructed as the absolute differences between percentage changes of positive and negative feedback investors, both defined in terms of return and volatility. We consider two specifications. In *the first specification*, the factors are constructed using the flows (both purchases and sales) *of positive and negative feedback investors*, defined on the basis of return and volatility. Each single portfolio is composed of *the percentage changes in both purchases and sales* of the investors belonging to the specific category. For example, the portfolio of negative return feedback investors (NRFI) is made of four components: a measure of dispersion constructed by using the purchases of the negative feedback investors identified on the basis of their sales, a measure of dispersion constructed by using the purchases of the negative feedback investors identified on the basis of their purchases, a measure of dispersion constructed by using the sales of the negative feedback investors identified on the basis of their sales and a measure of dispersion constructed by using the purchases of the negative feedback investors identified on the basis of their purchases. In *the second specification*, the four factors are constructed by using the purchases and sales separately considered, of *feedback investors*, regardless of the direction of their reaction (positive or negative feedback investors). We therefore have dispersion defined on the basis of purchases of return investors, dispersion defined on the basis of the purchases of volatility investors, sales of return investors, sales of volatility investors, net purchases (purchases minus sales) of return investors and net purchases of volatility investors. For example the portfolio of the purchases of return investors is made of four components of the measure of dispersion: the one based on the purchases of the negative return investors identified on the basis of their purchases, the one based on the purchases of the negative return investors identified on the basis of their sales, the one based on purchases of the positive return investors identified on the basis of their purchases and the one based on the purchases of the positive return investors identified on the basis of their sales. Loadings for each portfolio and portfolio weights are estimated via a principal component analysis performed on over-lapping 90 days windows through the sample period. A Dimson-Marsh correction using two days of leads and lags is applied to control for potential lead-lag effects due to asynchronous trading. The factor extraction and the estimation of the betas are updated each day in the sample, following the initial 90-day estimation period. Thus, betas are allowed to vary through time. In stage 2, we regress portfolio returns on betas for each day following the estimation period.

The table also reports the P-values of the tests which assess whether the means of the R^2 of the regressions with the dispersion of opinions-related as well as the dispersion of opinions-related factors are statistically different from the means of the R^2 estimated using only the standard asset pricing factors.

	Mean	**P**
I Specification (measure of dispersion of opinions based on *both* purchases and sales for positive and negative feedback investors, separately *considered*)		
Negative Return Investors	0.2459	0.004
Positive Return Investors.	0.2214	0.007
Negative Volatility Investors.	0.2169	0.530
Positive Volatility Investors.	0.2489	0.001
II Specification (measure of dispersion of opinions based on purchases and sales *separately considered* for *generic* feedback investors)		
Return Investors' Purchases	0.2133	0.010
Return Investors' Sales	0.2225	0.400
Volatility Investors' Purchases	0.2113	0.080
Volatility Investors' Sales	0.2061	0.370
Return Investors' Net Purchases	0.2289	0.003
Volatility Investors' Net Purchases.	0.2094	0.040

Table 7: Dispersion of opinions and consistent investors (residuals and dispersions of opinion)

The functional specification estimated is $Res_t = \alpha + \Sigma ß_k DB_{kt} + \varepsilon_t$, where Res_t are the residuals calculated from the second stage of a Fama-MacBeth procedure with 8 eight standard asset pricing factors (four extracted from past returns, four based on the investors' flows orthogonalized by regressing them on the first four factors). DB_{kt} are the dispersion of opinions. The standard asset pricing factors extracted from past returns are constructed using a principal component technique. We consider the regularly traded individual securities in the U.S. market. Loadings for each portfolio and portfolio weights are estimated via a principal component analysis performed on over-lapping 90 days windows through the sample period. The factors are extracted and loadings estimated using leading rolling windows. For the returns, we take the 560 stocks in the CRSP database that have been consecutively traded in the two-year period 1997-1998 with no missing observations. We then create 20 portfolios each containing 28 stocks, ranked by market capitalization. The standard asset pricing factors based on investors' flows are constructed by orthogonalizing the purchases and sales of the consistent investors on the first four factors constructed by using by using only past returns. We consider alternative specifications that differ depending on the type of flows we use to construct these four actors. We consider either the purchases or the sales of the investors identified in terms of positive (positive return or volatility investors) and negative (negative return or volatility investors) reactions, or the purchases and sales *separately considered* of the investors defined on in terms of the event they react to (return and volatility investors). Also the case when only the first 4 factors extracted from past returns is considered ("Return"). The dispersion of opinions-related factors are constructed using the same way of aggregating the transactions (purchases and sales) of different classes of consistent investors used to build the previously defined standard asset pricing factors. They are constructed as the absolute differences between percentage changes of positive and negative feedback investors, both defined in terms of return and volatility. The factors are constructed using the flows (both purchases and sales) *of positive and negative feedback investors*, defined on the basis of return and volatility. Each single portfolio is composed of *the percentage changes in both purchases and sales* of the investors belonging to the specific category. For example, the portfolio of negative return feedback investors (NRFI) is made of four components: a measure of dispersion constructed by using the purchases of the negative feedback investors identified on the basis of their sales, a measure of dispersion constructed by using the purchases of the negative feedback investors identified on the basis of their purchases, a measure of dispersion constructed by using the sales of the negative feedback investors identified on the basis of their sales and a measure of dispersion constructed by using the purchases of the negative feedback investors identified on the basis of their purchases. Loadings for each portfolio and portfolio weights are estimated via a principal component analysis performed on over-lapping 90 days windows through the sample period. A Dimson-Marsh correction using two days of leads and lags is applied to control for potential lead-lag effects due to asynchronous trading. The factor extraction and the estimation of the betas are updated each day in the sample, following the initial 90-day estimation period. Thus, betas are allowed to vary through time. In stage 2, we regress portfolio returns on betas for each day following the estimation period.

	Specifications									
	Return		Negative Return Investors		Positive Return Investors		Negative Volat. Investors		Positive Volat. Investors	
	Value	TStat	Value	TStat	Value	TStat	Value	TStat	Value	TStat
Constant	0.002	25.98	0.002	16.13	0.002	22.08	0.002	22.01	0.002	17.51
Return Investors' Purch.	-0.034	-2.90	-0.02	-0.36	-0.010	-0.11	-0.059	-0.68	-0.040	-0.54
Return Investors' Sales	0.045	0.38	0.27	1.50	0.277	2.11	0.286	2.13	0.262	1.64
Volat. Investors' Purch.	0.229	2.12	0.37	3.05	0.347	3.19	0.384	3.80	0.307	2.32
Volat. Investors' Sales	0.121	1.27	-0.19	-1.02	-0.063	-0.65	-0.080	-1.06	-0.016	-0.13
R Square	0.124		0.190		0.241		0.199		0.177	

	Specifications											
	Ret.Investor Purchases		Volat.Investor Purchases		Ret.Investor Sales		Volat.Investor Sales		Ret.Investor Net Purch..		Volat.Investor Net Purch.	
	Value	Tstat	Value	TStat	Value	TStat	Value	TStat	Value	TStat	Value	TStat
Constant	0.002	24.48	0.002	23.97	0.002	20.32	0.002	20.20	0.002	22.17	0.002	-0.04
Return Investors' Purch.	0.005	0.06	-0.056	-0.72	-0.048	-0.54	-0.034	-0.40	-0.03	-0.45	-0.115	0.23
Return Investors' Sales	0.123	0.91	0.170	1.33	0.231	1.50	0.307	2.16	0.35	2.12	0.270	0.15
Volat. Investors' Purch.	0.216	2.48	0.311	3.41	0.154	1.34	0.367	3.83	0.32	2.95	0.224	0.08
Volat. Investors' Sales	0.028	0.37	-0.079	-0.97	0.088	0.89	-0.043	-0.50	-0.06	-0.66	0.049	-0.04
R Square	0.144		0.149		0.107		0.246		0.210		0.146	