Heterogeneity of trade and stock returns. Evidence from index fund investors.

W. Goetzmann and M. Massa *
Yale University and Insead †

November 2000.

Abstract

We address the issue of how the heterogeneity of trade among investors affects stock returns. We model and test the relationship between dispersion of opinion, heterogeneity of trade and stock returns. The empirical investigation makes use of a two-year panel of more than 91 thousand individual accounts in an S&P 500 index mutual fund. We show that dispersion of opinion, proxied by the heterogeneity of trade among investors, explains part of the returns not accounted for by the fundamentals. We analytically and empirically show that the explanatory power of the dispersion of opinion increases at the very time when standard pricing models based on fundamentals fare worse.

*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 opinions among investors have been widely recognized as one of the main determinants of asset prices. However, scarce empirical evidence has been collected in order to support this thesis. In this paper we will lay-out a framework to analyze this issue by focusing on the relationship between investors' heterogeneity of trade, differences of opinions and asset prices in a context when the behavior of some investors is partially irrational. By using a dataset on investors' trades disaggregated at investor level we will empirically examine the relationship between investors' differences of opinions and asset returns.

Finance literature has extensively analyzed the link between differences of opinions among investors and asset prices. Differences of opinions have been justified in terms of either heterogeneity of beliefs or asymmetry of information. Williams (1977), incorporates heterogeneity of beliefs in the standard CAPM framework and shows that how they affect market returns. Detemple and Murthy (1994) prove that the equilibrium interest rate itself is a function of investors' beliefs, weighted according to the fraction of total wealth held they hold. Kraus and Smith (1989) argue that, even in the absence of new information about security payoffs, change in beliefs may move prices. The fact itself that investors are imperfectly informed about each other's endowments creates and reinforces uncertainty and preserve heterogeneity of opinion at equilibrium. This "market created risk" affects prices and equilibrium levels of returns.[1] In all these models investors' heterogeneity plays a key role in the price-formation process, however the restrictions directly testable in terms of investors' trade and holding positions are very few.

More recently Wang (1993) and He and Wang (1993) explicitly focus on the relationship between the asymmetry of information among investors and asset prices. In this case the dispersion in investors' is mainly due to differences in information sets rather than differences in opinion.

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

[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.

a direct estimation of it.

The first is that the link between heterogeneity of trade and dispersion of opinions about expected asset returns and risk is not directly evident. That is, is unclear how the process of aggregating investors' heterogenous holdings may affect the determination of stock returns. Indeed, Wang and Lo (2000) and Bossaerts, Plott and Zame (2000) have recently come to opposite conclusions regarding the relationship between investors' holdings and asset prices. Wang and Lo 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 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), Lakonishok and Maberly (1990) and more recently Grinblatt and Kellaharju (1999) use individual investor account data that allows analysis about 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 opinions, explicitly incorporating the possibility of irrational or biased beliefs about the market. In particular, we lay out a model where investors heterogeneity of trade, dispersion of opinions and returns are explicitly linked through directly testable restrictions.

We develop a model of an economy with informed, rational investors and uninformed, biased investors. In this setting, in a manner similar to Delong, Shleifer,

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

Summers and Waldmann (1990) the behavior of the biased investors, while irrational, has important pricing implications. An implication of the model is that the shifting beliefs of the uninformed investors are a source of risk, and become a key variable in the information set of informed investors. The influence of this variable on hedging, and thus on asset prices, is increasing in the dispersion of beliefs-represented in our model by the heterogeneity of trades between investors.

In the empirical analysis, we estimate the relationship between dispersion of opinions and returns. We use information on investors' purchases and sales into an index fund disaggregated at the individual level in order 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 opinions and asset returns. While most of the studies cited above use data on individual securities or funds, we focus on the relationship between stock prices and the dispersion of opinion about the stock market as a whole, rather than the relative investment prospects for individual securities.

The estimations support the model and show that dispersion of opinions explains part of the variance in returns not accounted for by the fundamentals.

2. The model

2.1. The economy

Assumption 1

We consider a standard Lucas' exchange economy with a single good (David 1987, Wang 1993, He and Wang 1993, Veronesi 1999). Dividends follow a geometric Brownian motion that can be represented as:

$$dD_t = \mu_D dt + \boldsymbol{\sigma}_D dz_D \tag{1}$$

where $\boldsymbol{\sigma}_D$ is diffusion of the real productivity process and μ_D its long term level. The level of dividends and their law of motion are known to *all* the investors in the market. Indeed, the existence of newsletters, brokers' reports, specialized press and research in general provides a widespread and cheap dissemination of information. This assumption allows us to explicitly focus on the asymmetry of information generated by investors' "irrational behavior".

Assumption 2

We assume two classes of investors: informed and uninformed. There are ω informed investors and $(1-\omega)$ uninformed investors. The informed investors represent the institutional investors: they have access to privileged source of information and therefore are fully informed. The uninformed, instead, represent the small investors: they are not fully informed and learn by observing prices. Both classes of investors behave rationally, by maximizing long term profits (Π_t). They solve the following problem:

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

where r is the riskless rate, D_t is the dividend, P_t is the price of the stock, X_t is the amount invested in stocks and ρ is the degree of risk aversion. We make the simplifying assumption that both informed and uninformed investors have the same degree of risk aversion $\rho = 1$. A different degree of risk aversion would not change the main results and would only make the exposition unnecessarily complicated. Q_t represents the excess return over the riskless asset. It can be represented as the return on a zero-wealth portfolio long one stock and fully financed by borrowing at the risk-free rate. In line with the standard literature (Sundaresan 1983, Wang 1993, He and Wang 1993, Naik 1997, Hau 1997) we assume a constant interest rate.[3] The informed investors's optimal demand of stocks is:

$$X_t^i = \frac{E[dR_t|\digamma_t^i]}{\rho E[dR_t^2|\digamma_t^i]} \tag{3}$$

that is, a direct function of the first and second moments of stock returns.

Assumption 3

The uninformed investors make systematic expectational errors (Ψ_t). In particular, uninformed investors' demand is:

$$X_t^u = \frac{E[dR_t|\digamma_t^u] + \Psi_t}{\rho E[dR_t^2|\digamma_t^u]} \tag{4}$$

These systematic errors may be interpreted as deviations from the optimal investment rule due to some social or behavioristic motives (Daniel, Hirshleifer and Subrahmanyam, 1997, ODean, 1998).[4] Therefore, the less informed investors

[3] This is equivalent to assuming that the interest rate is exogenously determined and that there is an infinitely elastic borrowing and lending facility.

[4] Unlike a standard model with idiosyncratic stochastic shocks to the endowments (Huang and Wang 1998, Marin and Rahi 1998), in this case uninformed investors are not fully aware of the fact that they are deviating from the rational behavior.

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 pseudosignals from technical analysts, stockbrokers, or economic consultants and irrationally believe that these signals carry information. Or in formulating their investment strategies, they may exhibit the fallacy of excessive subjective certainty that has been repeatedly demonstrated in experimental contexts since Alpert and Raiffa (1982)". More recently, Hau (1998) describes the behavior of similar traders in order to analyze the foreign exchange market.

This allows us to characterize uninformed investors' behavior as a mixture between a fully rational behavior (where Ψ_t is equal to zero and only $E[dR_t|\digamma_t]$ matters) and standard noise trading (where $E[dR_t|\digamma_t]$ is equal to zero and Ψ_t only matters). Ψ_t represents investors' "degree of overconfidence", that is they reactions to the general "climate of confidence" in the economy that makes them deviate from fully rational behavior, similar to the noise-trader factor in Delong et al (1990). Ψ_t adds persistence to investors' behavior and captures some of the "herding" or "momentum effects" identified by Barberis, Shleifer and Vishny (1998) and Lakonishok, Shleifer and Vishny (1993).

The very existence of Ψ_t makes the equilibrium not fully revealing for the less informed investors, as Ψ_t becomes an additional sources of uncertainty they have to filter out. It also provides institutional investors with an informational advantage - they know the uninformed investors' systematic error. Ψ_t is *exogenous* and behaves according to:

$$d\Psi_t = \Psi_t dt,$$

where Ψ_t can take two values: H and L. In the former case ($H > 0$) the investors' confidence is high, in the latter case ($L < 0$), confidence is low.

Therefore, this approach shares part of the salient features of the standard rational expectation models based on heterogeneous information (Wang, 1993, He and Wang, 1993) and part of the features of the behavioral models based on investors' "irrationality" (Delong, Shleifer, Summers and Waldmann,1990). In particular, in line with the latter models, small investors' irrational behavior can be due to their "following pseudosignals", such as "volume and price patterns, sentiment indices, forecasts of Wall Street gurus".

Assumption 4

The evolution of the confidence level (Ψ_t) moves between two states (H, L), according to a Poisson process with a transition probability matrix between time

t and time $t+dt$:

	H	**L**
H	$1-\lambda dt$	λdt
L	λdt	$1-\lambda dt$

The use of a Markov process that lets us model the transition between the two regimes is quite general. In fact, it allows the uninformed investors to develop expectations on the basis of past experience and derive a probability distribution. [5]

Assumption 5

One stock is traded. It is in positive net supply, with the supply normalized to be equal to 1.

2.2. Market equilibrium

To solve the model, we follow the standard approach. We first conjecture a linear pricing function. We then solve investors' inference and optimizing problems. Finally the market clearing condition allows us to back out the value of the coefficients for the linear pricing form we have conjectured. We conjecture a linear price such that:

$$P_t = p_0 + p_D D_t + p_\Psi \Psi_t + p_\Delta \Delta_t = p_0 + \Phi_t + p_\Psi \Psi_t + p_\Delta \Delta_t \tag{5}$$

where $\Delta_t = [E_t^u(\Psi_t|\digamma^u) - \Psi_t]$ is the learning error made by the uninformed investors and $\Phi_t = E_t[\int_t^\infty e^{-rs} D_s ds]$. 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 ($\Delta_t = 0$) and no uncertainty due to "extraneous" risk ($\Psi_t = 0$). This would be the case if the small investors behaved fully rationally and were not affected by the climate of confidence.

The stock price is therefore a linear function of the fundamental source of risk (D_t), of some "extraneous risk" due to small investors' irrational behavior (Ψ_t) and of additional risk induced by uninformed investors' learning process (Δ_t) (Wang 1993, He and Wang 1993, Naik 1997). The values of p_0, p_Ψ and p_Δ have to be defined in equilibrium and depend on investors' behavior, that is their degree of informativeness and "degree of irrationality".

[5] Another way of describing the behavior of Ψ_t, is by using a Poisson stochastic differential equation, such that:

$$dz_{\Psi_t} = (H-L)[1 - \frac{2(z_t - H)}{L-H}]dq_t$$

where $z_{\theta_0} \in \{H, L\}$, $H > L$ and q_t is a Poisson process with parameter $\lambda > 0$ (David 1997).

2.2.1. The inference problem

Informed investors know both D_t and Ψ_t, while the unformed investors only know D_t and by observing prices [6] they can infer the value of Ψ_t.

Theorem 1
The evolution of the posterior probability of the regime H is:

$$d\pi_H = \mu_{\pi_H} dt + \boldsymbol{\sigma}_{\pi_H} d\nu \tag{6}$$

where $d\nu. = \left(\frac{dP_t}{P_t} - \left[\pi_H H + (1-\pi_H)L\right] dt\right)$, $\mu_{\pi_H} = (1-2\pi_H)$ *and* $\boldsymbol{\sigma}_{\pi_H} = \frac{\pi_H(1-\pi_H)(H-L)}{\boldsymbol{\sigma}_D}$ (proof in Appendix 1).

2.2.2. Investors' stock demand

In order to get investors' demand defined in terms of their information set, we then substitute for investors estimations of the first and second moments of stock's excess return in equations 3 and 4

Theorem 2
Informed investor's demand is: $X_t^i = \frac{e_0 + e_\Psi \Psi_t + e_\Delta \Delta_t}{\rho V^i}$. *Uninformed investors' demand is:* $X_t^u = \frac{e_0 + e_{\widehat{\Psi}} \widehat{\Psi}_t + \Psi_t}{\rho V^u}$, *where* $V^i = V^u = p_D^2 \sigma_D^2$ (proof in Appendix 1). The level of confidence in the economy *unconsciously* affects small investors decision over and above their rational choice and acts as a sort of "focal point" that implicitly coordinates their investment decisions.

2.2.3. Equilibrium

In equilibrium the market clearing condition implies:

$$\omega X_t^i + (1-\omega) X_t^u = 1, \tag{7}$$

Substituting in equation 7 the optimal demand of stock by the two classes of investors and matching coefficients, we find that: $p_0 = -\frac{S\sigma_D^2}{r^3}$, $p_\Psi = \frac{1-\omega}{r-1}$ and $p_\Delta = -\frac{\sigma_\Psi (\omega-1)^2 (r-1)}{\omega[\omega - 1 + \sigma_D (r-1)^2]}$ (see Appendix 1).

[6] As a matter of fact they also abserve the dividends, but their information is identical to the one contained in the prices. Therefore, given that we have only one no-redundant signal two signals (the one contained in prices) and two sources of uncertainty, the equilibrium is never fully revealing to the uninformed investors.

2.3. Testable restrictions

By using equations 3 and 4, after some manipulations, we find the relationship between stock returns and investors' demands. In particular,

$$d(P_t) = d(\Phi_t) + \beta d(Disp_t), \tag{8}$$

where the values for α_0, α and β are provided in Appendix 2 and $Disp_t = [(1-\omega)X_t^u - \omega X_t^i] - [(1-\omega)X_{t-1}^u - \omega X_{t-1}^i]$ is investors' dispersion of opinions.

Therefore, the first testable restriction is a direct relationship between heterogeneity of investors' trade $(d(Disp_t))$ and asset returns. The size of the impact (γ) is related to relative size of each class of investors (ω) and to the volatility of the dividends (σ_D). [7]

It is worth noticing that the fact that the information content of the dispersion of opinions-related factors is mostly orthogonal to the one contained in price-based information. Thus, data on the purchases and sales of different types investors become useful for forecasting stock prices.[8]

Our solution provides a further basis for understanding the relationship between heterogeneity of trade and returns over time by focusing on the way the coefficients ("loadings") that relate stock price to such factors change. From now on, we will define the sources of uncertainty due to the fundamentals (Φ_t) as "fundamentals-related factors" and the sources of uncertainty due to investors' learning and differences of opinions $(Disp_t)$ as "dispersion of opinions-related factors". The loadings on the fundamentals are not affected by either over(under)confidence in the market or by investors' learning errors. Conversely the loadings on the dispersion of opinions (i.e. β) depend on both the information uncertainty and the degree of asymmetry of information in the market (ie. percentage of uninformed versus informed investors).

Let's see how this happens. In Graphs 1 we report simulated values of β for different levels of volatility of the dividends (σ_D) and fraction of uninformed investors in the market $(1-\omega)$. [9] It clearly appears that a reduction in the ability of the less informed investors to make correct inferences increases β. This lower

[7] It is worth noticing that we explicitly modelled Ψ_t as an underlying source of uncertainty related to investors' irrationality. However, we could have assumed it to be related to some other non fundamentals-related shock. For instance, shocks to the supply of stocks (e.g. SEOs). What we really are concerned about, is not so much the source of this additional shock, but the way learning uncertainty and differences of opinion about it affect stock returns. That is, how investors' way of interpreting it affects stock returns.

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

[9] We consider "plausible" values of b_D , that is the ones in general estimated and assuemed in the litearature (Campbell and Kyle 1993, Veronesi 199).

ability is due to higher fundamental uncertainty (σ_D) that makes the learning process more difficult. [10]. Also, the informational asymmetry in the market plays a role. The more "polarized" the market is, that is, more different is the number of informed investors as opposed to the number of uninformed investors in the market, the higher the loading on the dispersion of beliefs-related factor (β). Indeed, if there are very few informed investors (ω close to zero), the exogenous uncertainty (Ψ_t) induced by uninformed investors's behavior will be higher. On the other hand, if the uninformed investors are very few (ω close to 1), the risk premium due to informational asymmetry will be higher. [11] Given that the degree of informational asymmetry is a function of both the informational uncertainty of the less informed investors and the distribution of informed versus uninformed investors, it follows that the power to explain stock returns of the dispersion of opinions-related factors is a direct (positive) function of the degree of informational asymmetry in the market.

In terms of the empirical testing, this implies that we should be able to identify two "regimes" characterized in terms of market information asymmetry. We will now proceed to test these implications.

3. The data and definition of the variables

3.1. Data

Despite the increasing importance of index fund investing in the U.S. over the past two decades, there is relatively little information about the scale, activity and type of investor accounts that comprise an index fund. This is particularly bad as investors in the fund have explicitly chosen an index fund as opposed to a managed fund. Therefore, their behavior provides a very good proxy of overall market sentiment at a particular point in time.

Fidelity provided us with anonymous individual account activity in their 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 past five years, the fund has 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

[10] Indeed, dividends are the signals uninformd investors observe to filter out the value of Ψ_t.

[11] Wang (1993) already makes this point by noting 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 their informational disavantage of the less informed investors increses the premium they ask for informational asymmetry.

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

In particular, 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. 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]

3.2. 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 (since every trade is by definition two-sided). Massa and Goetzmann (1998) found that a measure of belief dispersion correlated well to the two proxies above, as well as to a measure of analyst disagreements, taken from timing newsletter forecasts. In that case, they used as an instrument for belief dispersion the sum of the absolute value of inflows and outflows from three Fidelity index funds. The logic of this measure is that when both inflows and outflows of the funds are high this indicates disagreement among investors about the prospects of the market. 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' heteterogeneity. Our dataset, by allowing us to separate the inflows and outflows by investor classes, provides us with a much more precise measure of uncertainty.

As a first rough measure of dispersion of opinions we use the absolute value of the difference between the trades (purchases and sales) of different classes of

[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.

investors. In particular, if we consider N classes of investors with $j = 1, ...i, ...N$, we can construct the following metric of investors' dispersion of trade:

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

This measure postulates a direct relationship between dispersion of holdings of the investors and dispersion of opinions.

A more refined measure exploits the structural relationship between investors' demand and underlying state variables. The model suggests that $X^i = f(D, \Psi, \widehat{\Psi})$ and $X^u = f(D, \widehat{\Psi})$. We can therefore use investors' holdings to back out the *perceived* underlying states. In particular, if we define the vector of investors' perceived states as $\mathbf{\Omega}_t$, with $\Omega^i_{f,t}$ as the entry for the *f-th* factor, investors' behavior can be described by the following system:

$$\mathbf{\Omega}_{t+1} = \mathbf{A}\mathbf{\Omega}_t + \mathbf{w}_t \tag{9}$$

$$X_t^j = \mathbf{C}\mathbf{\Omega}_t + v_t \tag{10}$$

where X_t^j are the holdings of the *jth* class of investors, while $\mathbf{w}_t$ and v_t are zero-mean normally distributed random variables with covariance matrices Q and R respectively. This is a linear dynamic system where only investors' holdings (X_t^j) are observed while the state and the noise variables are unobserved.

We estimate the system of equations 9 and 10 using a Kalman filter technique based on an expectation maximization algorithm (EM). This uses Shumway and Stoffer's (1982) methodology modified to allow for both A and C to be unobserved and to be simultaneously estimated. In particular, for each period, we estimate the conditional value of the states ($\widehat{\Omega}_t^j = E[\Omega_{t+1}]|\{X^j\}_1^t]$). This represents the expected value of the underlying states on the basis of the information contained in the holdings of the *jth* category of investors and available up to t. Daily time series of $\widehat{\Omega}_t^j$ are calculated for different classes of investors. Then Froboenius norms constructed as measures of dispersion of opinions. That is, for *each factor* ($\widehat{\Omega}^i_{f,t}$) we build:

$$Disp_t = \sum_f \sum_{j=1}^{N} Abs\left(\widehat{\Omega}^i_{f,t} - \widehat{\Omega}^j_{f,t}\right),$$

which represents the average difference between the perception of the *fth* factor between the *ith* and the *jth* class of investors.

How do we select the N classes of investors whose aggregate trades make up our X_t^i? The selection is done by grouping the individual accounts on the basis of the 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 problems in two ways. First, we explicitly consider an approach based on grouping investors in sample and tracing their strategies out of sample. This will be dealt with in the last section ("rational investors").

Alternatively, 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.

We therefore identify investors on the basis of five criteria. 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 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 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 determine the average value of these time series for the first 25 and the last 25 categories separately considered. The resulting time series provide the first two factors. 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.

3.3. Identification of "rational investors"

An alternative measure of dispersion of opinions relies on investigating the behavior of a subset of investors defined 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.

In particular, we define as positive feedback traders investors who purchased when the market rose and sold when the market fell 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 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. [13]

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

[13] 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. Thus we identify investors in terms of their reactions to changes in expectations about market risk. Both contrarians and momentum traders are therefore defined in terms of the reaction to the previous day returns/volatility.

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. This is consistent with Grinblatt and Kellaharju's findings that contrarians are more common in their sample than momentum investors. [14] 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%. [15]

A few words are needed with regards to what we actually mean by "feedback traders". This is simply one way of sorting the sample. Of course, given the institutional characteristics of the fund[16] we are concentrating on, 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. [17]

This test consists of examining 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 sub-periods: 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

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

[15] 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.

[16] That is a fund mostly used by investors with a long-terms investment horizon.

[17] Also, 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, 1998.), where some evidence is found that, on average, index fund investors reacted negatively to the previous day's market drop. In addition, index fund daily flows are correlated to the movement of the market in a manner suggesting that S&P 500 index fund investors may at times be salient to stock price formation.

reacting to either returns or volatility, where the median proportion measure is the dividing line between the two. [18] The values of the statistics (B) for the odds ratio test are significant for all classes except volatility momentum in sales and sales return and volatility contrarians. The results indicate that investor groups we identify typically display consistency over time. [19]

Once the times series of aggregate purchases and sales of investors' classes have been constructed, we 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 opinions. In general, we consider two specifications.

In the first specification, the factors are constructed using the purchases and sales of positive and negative feedback investors, defined on the basis of return and volatility. Each single factor is composed of *both* purchases and sales of the investors belonging to the specific category. For example, the portfolio of negative return feedback investors ($NRFI$) 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

[18] 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.

[19] 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.

purchases of return investors is made of four components of the measure of dispersion: the component based on the purchases of the negative return investors identified on the basis of their purchases, the component based on the purchases of the negative return investors identified on the basis of their sales, the component based on purchases of the positive return investors identified on the basis of their purchases and the component based on the purchases of the positive return investors identified on the basis of their sales.

One natural question is whether our measure of dispersion of opinions correlates with the standard measures of market uncertainty (implied volatility) and dispersion of opinion (open interest and trading volume). To test this, we regress the market polarization 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 market polarization and the standard proxies for dispersion of opinions (open interest and trading volume). No correlation, however, is found between market polarization and implied volatility. This fits with Prabhala (1998) who pins down open interest as a measure of dispersion of opinion and implied volatility as a measure of market uncertainty, orthogonal one to the other.

4. Empirical estimation

The next step is to test whether there is a correlation between the explanatory power of standard asset pricing models and our measures of dispersion of opinion. This may be tested in different ways. We adopt three approaches.

First, we consider the relationship between stock returns and dispersion of opinions using a standard asset pricing framework. The model predicts that the relative explanatory power of the fundamentals and the dispersion of opinions depend on the overall market learning uncertainty. An increase in learning uncertainty raises the explanatory power of the weight of the dispersion of opinions-related factors and reduces the weight of the fundamentals-related factors. This can be tested by considering the incremental explanatory power of the dispersion of opinions-related factors after having accounted for the explanatory power of the fundamentals-related factors.

However, equation 8 also suggests that the relation between stock returns and dispersion of opinions may be better characterized in terms of the existence of two different *regimes*. In the first regime, characterized by a high degree of information asymmetry in the market, the role of the dispersion of opinion is high. In the second regime, characterized by a low degree of information asymmetry, the importance of the dispersion of opinion drops and fundamentals explain most of the stock returns. We can therefore use a Markov-switching model to test this hypothesis.

Finally, we test whether measures of dispersion of opinions based on their behavior of "rational investors" - i.e. that display some consistent rational patterns - have explanatory power.

4.1. Dispersion of opinion and asset prices

Given that we want to assess the incremental explanatory power of the dispersion of opinions-related factors, we first orthogonalize these factors, by regressing them on the fundamentals-related ones by estimating the auxiliary regression:

$$F_{D,t} = \theta + \delta F_{r,t} + \varepsilon_t. \tag{11}$$

$F_{r,t}$ are the fundamentals-related factors extracted by using stock returns and $F_{D,t}$ are the dispersion of opinion-related factors. Fundamentals-related 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.[20] The factor extraction is performed daily, following the initial 90-day estimation period. In particular, we estimate loadings for each portfolio and portfolio weights via a principal component analysis performed on overlapping 90 days windows through the sample period.

In this way we capture the relevant factors driving the cross-section of returns in the preceding 90-day window. If, for example, these true factors were a rotation of the Fama-French factors they should be captured in the first stage. If they are well estimated, then we would expect their loadings to explain cross-sectional dispersion in returns in the following period. Then, we test whether the dispersion of beliefs-related factors have any additional incremental explanatory power to the one already latent in the fundamentals-related market returns. This is done by estimating:

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

where $\varepsilon_{i,t}$ are the residuals from the equation 11 and $R_{i,t}$ are the returns on portfolios of stocks. In particular out of the selected sample, we create 20 portfolios each containing 28 stocks, ranked by market capitalization.

Equation 12 is estimated by using a standard Fama-MacBeth [FM] two-stage time-series cross-section test, applied to daily returns. We applied it to rolling intervals and daily updated betas. Given that we need a 90-day rolling window

[20]The reason why 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.

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. Given that we are dealing with daily data with potential lead-lag effects due to asynchronous trading, 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 $\overline{r}_t^2$. [21]

If our working 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, [22] allows us the gauge the role played by dispersion of opinion on asset prices. The results, reported in Tables 3, strongly support our hypothesis, displaying a significant additional explanatory power provided by the factors based on investors' dispersions of holdings. This holds for all the specifications that have been considered. The values of the $\overline{r}_t^2$ are very similar, regardless of the specification (either based on perceived states or based on actual flows) and of the criterion used to classify investors (by trading frequency, by dispersion of holdings, ...). [23] It is worth noting that the explanatory power of the dispersion of opinions-related factors is not very dissimilar from the one of the fundamentals-related factors (on average 9.5-10.5% as opposed to the 12.6%).

Furthermore, in order to explicit test the statistical significance of the incremental explanatory power, we estimate:

$$R_{i,t} = \alpha_i + \beta_i \varepsilon_t + \gamma_i F_{r,t} + \eta_{i,t}. \tag{13}$$

If the dispersion of opinion-related factors loads on a different source of uncertainty that the fundamentals-related ones, we expect them to increase the explanatory power in equation 13. Also we can explicitly test for the significance of the incremental explanatory power by comparing the $\overline{r}_t^2$ of the specification with both sets of factors to the specification with only the fundamentals-related ones. A *t-test* for distribution with unequal variances is therefore performed. The

[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.

[22] We know that, while fundamentals are autocorrelated over time, the dispersion of opinions 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 fundamentals-related factors and scarcely related to the didpersion of beliefs-related factors.

[23] This is important as it provides a good robustness check that the results are not biased by the implicit endogenity 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 endogenity bias.

results are reported in Table 4. They support the hypothesis that the dispersion of opinion-related factors provide a significant improvement of the explanatory power.

4.2. Dispersion of opinion and regimes

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 degree of information asymmetry in the market (H) and one characterized by low degree of 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 8 as:

$$R_{H,t} = \alpha_H + \beta_H F_t + \gamma_H Disp_t + \varepsilon_{H,t} \tag{14}$$

$$R_{L,t} = \alpha_L + \beta_L F_t + \gamma_L Disp_t + \varepsilon_{L,t}, \tag{15}$$

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. We estimate two specifications: one that uses investors' purchases and sales as measure of dispersion of opinion and one that uses investors' perceived states. F_t proxies for the "fundamentals", that is all the components of returns not related to dispersion of opinion. They are defined using some standard information variables (Ferson and Harvey, 1999) such as dividend yield, yield on long term corporate bonds (AAA quality), yield on junk bonds, yield on the Treasury Bills. 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 fundamentals. 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 is affected by the fundamentals (D_t), overall volume would provide an indirect proxy for it. This suggests a role for overall volume as related mostly to the fundamentals that is different form casual identification of overall volume with dispersion of opinion. We estimate two alternative specifications: one based on flows (purchases and sales) and one based on the "perceived states", as defined before. 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.

As 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. [24] 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 tends to offset each other, the aggregate impact of the dispersion of opinion-related factors differs in the two regimes: strongly positive in the second regime, on average null or slightly negative in the first one. This supports the working hypothesis. Dispersion of beliefs-related factors affect stock returns in a different way, depending on the degree of informational asymmetry in the market.

4.3. Dispersion of opinion and rational investors

The previous results suggest that investors' trade used to proxy for dispersion of opinion has incremental explanatory power with respect to fundamental-related 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 "rational ones" defined above) as a measure of dispersion of opinion. This has the merit of giving an economic intuition to the identification of

[24] When the additional controls of lagged past returns and overall trading volume are added, significance drops and in a few acses disapeprs. On average, however, they are always strongly signiicant and positive.

the dispersion of opinion-related factors which were before otherwise only defined on the basis of statistical techniques. We are also able to trace out-of-sample the behavior of investors defined in sample on the basis of rational strategies. We start from equation

$$R_{i,t} = \alpha_i + \beta_i F_{r,t} + \gamma F_{D,t} + \varepsilon_{i,t}, \tag{16}$$

where $F_{r,t}$ are the fundamental-related factors and $F_{D,t}$ are the dispersion of opinion-related factors. We consider eight fundamentals related 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 consider 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 fundamental-related factors.

Unlike equation 12 in the previous section, here we run a direct horse-race between fundamentals and dispersion of opinion. We also impose a more demanding condition by controlling for past investors' flows. If these flows have explanatory power, given the independence over time of investors' learning errors, they should capture fundamentals-related uncertainty not accounted for by the factors which have been extracted from past returns.

As before, our goal is to test whether adding our measures of dispersion of opinion increases the explanatory power. Indeed, if the opinion condition the explanatory power of the standard asset pricing model, we expect to find that our measure of market polarization adds significant 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 16. [25]. Then, we regress these residuals on our measure of belief dispersion as:

$$\varepsilon_{i,t}, = \alpha + \beta F_{D,t} + \eta_t, \tag{17}$$

where different specifications of the dispersion of opinion-related factors ($F_{D,t}$). [26] This allows us to see whether the dispersion of opinion explains the residuals. Furthermore, 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 fundamentals-related factors have a relatively lower power, we would expect: $\beta < 0$.

Indeed, 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. [27]

[25] 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)

[26] 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 4 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 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.

[27] Alternatively we consider the correlation between the explanatory power of the aforesaid standard FM model and our measure of dispersion of beliefs. In particular we estimate:

5. Conclusion

In this paper, we developed a model consistent with previous analysis of limited-rational investors, and we used this model to study the way heterogeneity of trade among investors affects stock returns. This framework allows us to examine the interaction between dispersion of opinion and asset prices.

Focusing on the pricing of the aggregate U.S. equity market, we then tested its empirical implications by using an unique dataset on index funds' investors. We showed that dispersion of opinion, proxied by the heterogeneity of trade among investors, explains part of the returns not accounted for by the fundamentals. In particular, we showed analytically and empirically that the explanatory power of the dispersion of opinion increases at the very time when pricing models based on fundamentals fare worse.

These results suggest a new 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 fundamentals.

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

$$\overline{r}_t^2 = \alpha + \beta F_{D,t} + \eta_t,$$

where $\overline{r}_t^2$ is the average adjusted $r_t^2$of the cross-section regression of the stock eturns on the on the pre-estimated betas. The estimations are carried out using GMM, with Newey-West correction on the variance-covariance matrix based on a five lag autocorrelation structure. This is done in order to capture weekly effects. Instruments are used to correct for errors-in-variables and measurement problems. Different specifications are considered, where the measures of dispersion of beliefs based on purchases and sales are jointly considered and the cases where they are separately used. Given that the results agree with the ones in the text, we do not report them. They are available upon request from the authors.

6. Appendix 1

Proof of Theorem 1

In general, if we assume that (θ, ξ) is a two-dimensional partially observable random process where ξ is the observable component, θ is the unobservable component and $\mathbf{E}$ is the set of possible values that the unobservable component (θ) can take. In particular, let's assume that the unobservable component follows:

$$d\xi_t = A_t(\theta_t, \xi)dt + B_t(\xi)dW_t$$

where W_t is a Wiener process.. From Liptser and Shiryayev (pag. 333) we know that the posterior probability of the state $\beta \in \mathbf{E}$ is:

$$\pi_\beta(t) = p_\beta(t) + \int_0^t \Re\pi_\beta(u)du + \int_0^t \pi_\beta(u)\frac{A_u(\beta,\xi) - \overline{A}_u(\xi)}{B_u(\xi)}d\overline{W}_u \tag{18}$$

where: $\Re\pi_\beta(u) = \sum_{\gamma\in\mathbf{E}} \lambda_{\gamma\beta}(u)\pi_\gamma(u)$, $\overline{A}_u(\xi) = \sum_{\gamma\in\mathbf{E}} A_u(\gamma,\xi)\pi_\gamma(u)$ and $\overline{W} = (\overline{W}_t, \Im_t)$ is a Wiener process with: $\overline{W}_t = \int_0^t \frac{d\xi_u - \overline{A}_u(\xi)}{B_u(\xi)}$. Here $\Im_t$ is the information set available at time t. In our case, the unobservable component (θ) can take values H and L (that is $E = [H, B]$). The observable components are D (the dividends) and p (the price level).

In our case, investors observe a signal (Ψ) which can be backed out from prices and try to infer the value of θ. Applying eq.18 to our case, we have:

$$\begin{aligned} d\pi_H &= (1-2\pi_H)\lambda dt + \frac{\pi_H(1-\pi_H)(H-L)}{\sigma_\Psi^2}\left(\frac{d\Psi_t}{\Psi_t} - [\pi_H H + (1-\pi_H)L]\right)dt = \\ &= \mu_{\pi_H}dt + \boldsymbol{\sigma}_{\pi_H}d\nu, \end{aligned}$$

where:

$$\begin{aligned} \mu_{\pi_H} &= (1-2\pi_H)\lambda,\ \boldsymbol{\sigma}_{\pi_H} = \frac{\pi_H(1-\pi_H)(H-L)}{\boldsymbol{\sigma}_\Psi} \\ \text{and } d\nu. &= \left(\frac{d\Psi_t}{\Psi_t} - [\pi_H H + (1-\pi_H)L]\, dt\right) \end{aligned}$$

Proof of Theorem 2

The price of the stock is conjectured to be:

$$P_t = p_0 + p_D D_t + p_\Psi \Psi_t + p_\Delta \Delta_t$$

where $\Delta_t = \widehat{\Psi}_t - \Psi_t$ and $p_D = \frac{1}{r}$. Following Campbell and Kyle (1993) and Wang (1993), we define the fundamental value of the stock as $\Phi_t = E_t[\int_t^\infty e^{-rs}D_s ds]$.

$$P_t = p_0 + \Phi_t + p_\Psi \Psi_t + p_\Delta \Delta_t.$$

Prices change according to the following law of motions:

$$dP = [p_D D_t + p_\Psi \Psi_t + p_\Delta \Delta_t]dt + b_P p_D dz$$

and the "excess-return" moves according to:

$$\begin{aligned} dR_t|F^i &= (D_t - rP_t|F^i)dt + dP_t|F^i = \\ & [e_0 + e_\Psi \Psi_t + e_\Delta \Delta_t]dt + b_P p_D dz, \end{aligned}$$

for the informed investors and

$$\begin{aligned} dR_t|F^u &= (D_t - rP_t|F^u)dt + dP_t|F^u = \\ & [e_0 + e_\Psi \widehat{\Psi}_t]dt + b_P p_D dz, \end{aligned}$$

for the uninformed investors. Here $e_0 = -rp_0$ and $e_\Psi = -p_{\Psi.}(r-1)$.Also, following David (1997) and Liptser and Shiryayev, we have that:

$$d\widehat{\Psi}_t = \widehat{\Psi}_t dt + \boldsymbol{\sigma}_{D_t} d\nu \tag{19}$$

where $\widehat{\Psi}_t = [\pi_H H + (1-\pi_H)L]$. This implies that $e_\Delta = -p_\Delta[r - (1 + \frac{p_\Psi}{\sigma_\Psi})]\Delta_t$, where $\Delta_t = \widehat{\Psi}_t - \Psi_t$. Therefore, we can define: $E[dR_t|F_t^i] = e_0 + e_\Psi \Psi_t + e_{\widehat{\Psi}} \widehat{\Psi}_t$, $E[dR_t|F_t^u] = e_0 + (e_\Psi + e_{\widehat{\Psi}})\widehat{\Psi}_t$, $E[dR_t^2|F_t^i] = E[dR_t^2|F_t^u] = p_D^2 \sigma_D^2$.

Equilibrium

In equilibrium, the market clearing condition requires that:

$$\omega X_t^i + (1-\omega)X_t^u = S,$$

that is,

$$\begin{aligned} \frac{\omega}{\rho V} e_\Psi + \frac{(1-\omega)}{\rho V} e_\Psi &= S \\ \frac{\omega}{\rho V} e_\Psi + \frac{(1-\omega)}{\rho V}(1 + e_\Psi) &= 0 \\ \frac{\omega}{\rho V} e_\Delta + \frac{(1-\omega)}{\rho V} e_\Psi &= 0. \end{aligned}$$

Substituting in equation 7 the optimal demand of stock by the two classes of investors and matching coefficients, we find that: $p_0 = -\frac{S\sigma_D^2}{r^3}$, $p_\Psi = \frac{1-\omega}{r-1}$ and $p_\Delta = -\frac{\sigma_\Psi(\omega-1)^2(r-1)}{\omega[\omega - 1 + \sigma_D(r-1)^2]}$.

7. Appendix 2

By using equations 3 and 4, we find that:

$$P_t = \alpha_0 + \Phi_t + \alpha[(1-\omega)X_t^u + \omega X_t^i] + \beta[(1-\omega)X_t^u - \omega X_t^i], \tag{20}$$

where, defining $\alpha_0 = a_0 + (\frac{dc}{f} - fa)\frac{a_1}{(b-e)}$; $\alpha_1 = \frac{a_1 f}{(b-e)}$ $\beta = (b_1 - \frac{a_1 c}{f(b-e)})$,we have that: $a_0 = \frac{\sigma_D^3 S(2\omega-1)(r-1)}{2\omega[-1+\omega+\sigma_D(r-1)^2]r^2}$, $a_1 = \frac{(\omega-1)-2(\omega-1)\omega}{2\omega[-1+\omega+\sigma_D(r-1)^2](t-1)}$ and $b_1 = \frac{\sigma_D^3(r-1)}{2\omega[-1+\omega+\sigma_D(r-1)^2]r^2}$.

Given that we assume that the net supply of stocks is constant in the short run and equal to S, we can write:

$$P_t = \alpha_0 + \Phi_t + \alpha S + \beta[(1-\omega)X_t^u - \omega X_t^i]. \tag{21}$$

Therefore, we can write the law of motion of stock returns as:

$$dP_t = d\Phi_t + \beta[(1-\omega)dX_t^u - \omega dX_t^i] = d\Phi_t + \beta Disp_t. \tag{22}$$

where $Disp_t = [(1-\omega)dX_t^u - \omega dX_t^i]$.

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

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

[5] 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.

[6] 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.

[7] Bussey J. 1998, " Volatility Timing in mutual funds: evidence from daily returns", N.Y.U. Stern Business School, Mimeo.

[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] Chevalier J. and G. Ellison, August 1995, "Risk taking by mutual funds as a response to incentives", NBER Working Paper #5234.

[13] Conrad J. and G. Kaul, 1998, "An anatomy of trading strategies", The Review of Financial Studies, 11(3), 489-519.

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

[15] David A., 1997, "Fluctuating Confidence in Stock Markets: Implications for Returns and Volatility", Journal of Financial and Quantitative Analysis; 32(4), 427-62.

[16] 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.

[17] 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.

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

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

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

[21] 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.

[22] 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 .

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

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

[25] 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.

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

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

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

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

[30] 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.

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

[32] Jeegadesh N. and S. Titman, 1993, "Returns to buying winners and losers, implications for stock market efficiency", Journal of Finance, 48, 65-92

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

[34] 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.

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

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

[37] Lakonishok J., Shleifer A. and R. W. Vishny, 1993, "Contrarian Investment, Extrapolation, and Risk" University of Chicago Center for the Study of the Economy and the State Working Paper: 84.

[38] 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.

[39] 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.

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

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

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

[43] Shumway, R.H. and Stoffer D.S., "An approach to time series smoothing and forecasting using the EM algorithm", Journal of Time Series Analysis, 3(4), 253-264.

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

[45] Stulz R. M., 1997, "Do foreign investors destabilize stock markets? The Korean experience in 1997", Mimeo, Charles A. Dice Center for Research in Financial Economics.

[46] Stulz R. M., 1999, "International portfolio flows and security markets", Mimeo, The Ohio State University.

[47] Veronesi, P., 1999, "Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model", Review-of-Financial-Studies; 12(5), 975-1007..

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

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

[50] Zheng, L. 1998, "Who Moves the Market," Working paper, University of Michigan Business School.

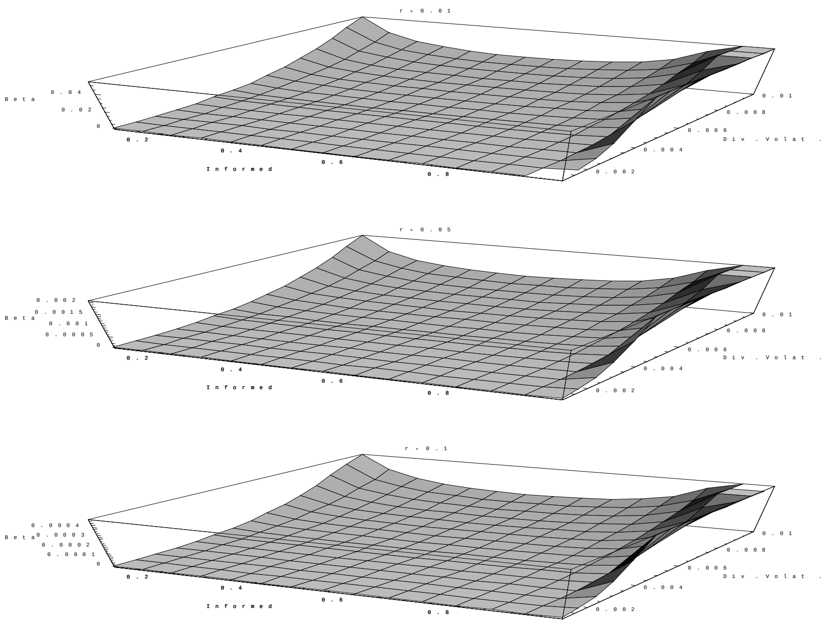

Figure 1: Values of β for alternative values of ω, r and b_D ($b_\Psi = 1$).

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: Typology of "Rational Investors"

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. Rational 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 rational 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	$\alpha>0.1$	1.19	0.10	1.04	0.12	1.06	0.11	1.51	0.14	1.08	0.11
	$0.5>\alpha>0.1$	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	$0.5>\alpha>0.1$	20.87	11.99	17.38	10.76	19.55	9.58	20.70	11.11	18.22	10.91
	$\alpha>0.1$	2.90	0.34	2.15	0.24	2.91	0.25	3.21	0.55	2.36	0.27
VPFI	$\alpha>0.1$	1.84	0.14	1.28	0.08	1.20	0.15	1.42	0.18	1.38	0.09
	$0.5>\alpha>0.1$	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	$0.5>\alpha>0.1$	17.71	10.41	15.94	9.70	17.29	8.54	19.00	10.83	16.41	9.79
	$\alpha>0.1$	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	$\alpha>0.1$	10.40	0.89	12.38	1.26	11.07	2.29	17.16	1.49	11.99	1.24
	$0.5>\alpha>0.1$	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	$0.5>\alpha>0.1$	27.11	13.78	23.85	15.28	21.37	11.07	21.64	11.19	24.40	14.60
	$\alpha>0.1$	16.00	1.60	12.91	1.99	14.12	1.91	14.18	3.73	13.71	1.94
VPFI	$\alpha>0.1$	16.44	1.42	14.93	0.88	14.89	1.91	12.69	2.99	15.21	1.11
	$0.5>\alpha>0.1$	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	$0.5>\alpha>0.1$	28.62	12.36	30.49	13.37	29.77	9.54	37.31	8.96	30.21	12.82
	$\alpha>0.1$	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 daily cross-sectional of 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 fundamentals-related factors. In the case of "Perceived States", the dispersion of opinions-related factors are constructed in the same way, but instead of purchases and sales, we considered the underlying perceived states estimated by using a Kalman Filter technique. The *fundamentals-related 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.

	Flows-based Dispersion of Opinions	
	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

	Perceived States-based Dispersion of Opinions	
	I Specification	**II Specification**
Full Set of Factors. Classification based on :		
Holding Dispersion	0.101	0.090
Running Balance	0.105	0.102
Average Holdings	0.102	0.094
Number of Transactions	0.102	0.091
Portfolio Turnover	0.103	0.102

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

The table reports the means of the R^2 from the daily cross-sectional of the second stage of a Fama-MacBeth procedure with 8 factors: four fundamental-related factors and four dispersion of opinions-related factors. *The fundamental-related 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 (fundamentals-related factors). In the case of "Perceived States", the dispersion of opinions-related factors are constructed in the same way, but instead of purchases and sales, we considered the underlying perceived states estimated by using a Kalman Filter technique. 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 fundamentals-related factors (4 factors) and the case based on both fundamental-related 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 fundamentals-related factors.

	Flows-based Dispersion of Opinions			
	I Specification		**II Specification**	
	Mean	**P Value**	**Mean**	**P Value**
Fundamentals-Related 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

	Perceived States-based Dispersion of Opinions			
	I Specification		**II Specification**	
Full Set of Factors.	**Mean**	**P Value**	**Mean**	**P Value**
Classification based on :				
Holding Dispersion	0.175	0.0001	0.177	0.0001
Running Balance	0.172	0.0001	0.169	0.0001
Average Holdings	0.163	0.0010	0.171	0.0001
Number of Transactions	0.157	0.0020	0.181	0.0001
Portfolio Turnover	0.175	0.0001	0.170	0.0001

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 ("Flows"). The resulting time series provides the first factor ($DB_{1,t}$). The other factor is calculated analogously by using the sales ($DB_{2,t}$). 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 ("Perceived States"). The dispersion of opinions-related factors are then orthogonalized by regressing them on the first four factors (fundamentals-related factors). In the case of "Perceived States", the dispersion of opinions-related factors are constructed in the same way, but instead of purchases and sales, we considered the underlying perceived states estimated by using a Kalman Filter technique. We consider alternative specifications that differ, depending on whether lagged returns or contemporaneous trading volume on the S&P500 are 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

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

	Flows-based Dispersion of Opinions							
	$ß_1$				**$ß_2$**			
	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

	Perceived States-based Dispersion of Opinions							
	$ß_1$				**$ß_2$**			
	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	0.01	0.09	1.23	2.76	-0.81	-6.70	1.04	4.56
Running Balance	0.02	0.23	1.19	2.34	-0.57	-6.02	0.95	4.77
Average Holdings	0.02	0.16	1.37	2.77	-0.69	-5.99	1.04	4.68
Number of Transactions	0.00	0.01	1.24	2.79	-0.74	-5.86	1.04	4.41
Portfolio Turnover	0.00	-0.03	1.30	2.91	-0.75	-6.28	1.03	4.53

II Specification

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

	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	0.01	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

	Perceived States-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	0.01	0.49	0.10	2.40	-0.07	-5.34	0.08	3.29
Running Balance	0.01	0.51	0.11	2.56	-0.05	-4.57	0.06	3.13
Average Holdings	0.01	0.62	0.10	2.37	-0.06	-4.49	0.08	3.37
Number of Transactions	0.01	0.49	0.10	2.41	-0.06	-4.42	0.08	3.19
Portfolio Turnover	0.00	0.42	0.10	2.48	-0.06	-4.81	0.08	3.32

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

	Perceived States-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	0.01	0.09	1.23	3.01	-0.82	-6.46	0.77	3.37
Running Balance	0.01	0.12	1.20	2.36	-0.61	-6.31	0.94	4.56
Average Holdings	0.02	0.20	1.28	3.06	-0.68	-5.60	0.75	3.43
Number of Transactions	0.01	0.12	1.01	2.02	-0.70	-6.18	1.15	4.24
Portfolio Turnover	0.00	0.00	1.27	3.19	-0.75	-5.95	0.73	3.15

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

The table reports the means of the R^2 from the daily cross-sectional of the second stage of a Fama-MacBeth procedure with 12 factors: eight fundamentals-related 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 fundamental-related 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 rational 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 fundamentals-related 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 rational investors
(residuals and dispersions of opinions)

The functional specification estimated is $Res_t = \alpha + \Sigma ß_k DB_{kt} + \varepsilon_t$, where Res_t are the residuals calculated from the daily cross-sectional of the second stage of a Fama-MacBeth procedure with 8 eight fundamentals-related 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 fundamental-related 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 fundamentals-related factors based on investors' flows are constructed by orthogonalizing the purchases and sales of the rational 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 *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 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 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 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	