Laboratoire
de Recherche
en Gestion
& Economie

Working Paper

Working Paper 2012-01

Portfolio diversification dynamics of individual investors: a new measure of investor sentiment

Patrick Roger

January 2012

Université de Strasbourg
Pôle Européen de Gestion et d'Economie
61 avenue de la Forêt Noire
67085 Strasbourg Cedex
http://ifs.unistra.fr/large

IFS
Institut de Finance
de Strasbourg

Portfolio diversification dynamics of individual investors: a new measure of investor sentiment

Patrick ROGER[1]

LARGE Research Center

EM Strasbourg Business School

University of Strasbourg

November 10, 2011

[1]We thank Marie-Hélène Broihanne and Maxime Merli for stimulating discussions on earlier versions of the paper, Laurent Deville, Gunter Franke, Jens Jackwerth for helpful suggestions, Malcolm Baker and Jeffrey Wurgler for the availability of their sentiment data (http://people.stern.nyu.edu/jwurgler) and Tristan Roger for research assistance. The financial support of CCR Asset Management is gratefully acknowledged.

Abstract

We build a new measure of investor sentiment only based on changes in diversification levels of individual investors' portfolios. The dynamics of the number of different stocks in portfolios is modelized as a Markov chain. We measure investor sentiment as the area above the cumulative distribution of the steady-state equilibrium of diversification levels.

We apply this model to a large sample of more than 80000 individual investors over the period 1999-2006. We first show that our index is significantly correlated to the French consumer sentiment index, to the Baker and Wurgler sentiment indices and to the buy-sell imbalance index, despite the fact we use neither prices or returns on stocks nor transaction volumes or even the identification of stocks bought or sold by the investors. Following the two-step methodology of Baker and Wurgler (2006), we show that our measure outperforms the others in predicting returns of a long-short portfolio based on size.

Keywords: Investor sentiment, retail investors, markov chains

JEL classification: G11, G14

Introduction

In this paper, we build a new market sentiment index (henceforth MSI) based on changes in portfolio diversification by individual investors. Our starting point is the well documented fact that retail investors hold underdiversified portfolios[1]. We postulate that an underdiversified investor who buys a new stock, that is a stock not already held, signals her optimism about future prices or possibly more generally about future economic conditions. We describe the dynamics of the number N of *different* stocks in portfolios as a Markov chain. By assuming that one-period transitions of this process between two dates $t-1$ and t are stable over time, we calculate the steady-state equilibrium of the Markov chain. It gives the proportions of investors holding one stock, two stocks, and so on, in the long-run, if the sentiment revealed by diversification changes were staying the same. Our date-t sentiment index is then measured by the area above the long-run cumulative distribution function of the number of stocks in portfolios. The intuition is that when usually under-diversified investors increase the number of stocks in their portfolio, they act on an optimistic view of future prices/returns.

We perform the same calculations at all dates to get a time-series of the sentiment index. As our empirical analysis focuses on the power of a sentiment index to explain future returns, we build an orthogonalized version of MSI by taking the residuals of the regression of MSI on the Fama-French (1992) factors and the Carhart (1997) momentum factor.

Our sentiment measure differs from the ones based on buy-sell imbalances[2] at least for two reasons. First, we only focus on purchases and sales that increase or decrease diversification, that is trades changing the number of different stocks in the portfolio. Second, we do not consider the volume of trade in order to disentangle the effect of sentiment (optimism/pessimism) from a potential demand/supply

[1] Lease *et al.* (1974) and Blume and Friend (1975) were the first to highlight the portfolio underdiversification of retail investors, followed by Kelly (1995). More recently, a number of empirical studies (Odean 1999, Mitton and Vorkink 2007, Kumar 2007, Goetzman and Kumar 2008) obtained the same results on large samples of U.S individual investors. Calvet *et al.*(2007), Broihanne *et al.* (2011) get similar patterns on the Swedish and French markets.

[2] For example Kumar and Lee(2006), Schmitz *et al.* (2007), Andrade *et al.* (2008), Hvidkjaer (2008), Barber and Odean (2008), Kaniel *et al.* (2008), Barber *et al.* (2009) use buy-sell imbalance measures either to measure sentiment or to analyze correlated trading among individual investors.

effect of retail trades.

Though we try to isolate a measure of expectations, our MSI also differs from survey-based measures like the index of consumer sentiment (ICS) of the university of Michigan, the investor sentiment survey of the American Association of Individual Investors or the French sentiment index, because survey-based indices measure what people think about future financial and economic conditions but do not control for what people actually do.

The third approach to build a sentiment index is top-down, that is based on macroeconomic variables. For example, Baker and Wurgler (2006) define their sentiment index as a linear combination of six variables, namely the closed-end fund discount, the logarithm of the NYSE share turnover ratio (detrended by the 5-year moving average), the number of IPOs, the average first-day return on IPOs, the share of equity issues in total equity and debt issues and finally the dividend premium defined as the log difference of the average market-to-book ratios of dividend payers and non payers. The linear combination of these indicators is chosen as the first principal component of a PCA of the six variables. Baker and Wurgler (2006) also define an orthogonalized version of their sentiment index by first regressing each of the six variables on growth in the industrial index, growth in consumer durables, non durables and services and a dummy variable for NBER recessions. They then consider the inputs of the PCA as the residuals of these regressions to define the orthogonalized version of their sentiment index.

In order to validate our measure, we build a time-series of MSI using trading records and portfolios of a large sample of 87,373 French individual investors over the period 1999-2006. We compare our MSI to other sentiment indices picked in the abovementioned categories, namely the French consumer sentiment index published by National Institute of Statistics and Economic Studies[3], the two Baker-Wurgler indices[4] and the Buy-Sell imbalance measure (BSI).

[3] INSEE: Institut National de la Statistique et des Etudes Economiques.

[4] All these indices are briefly described in the Appendix and references providing the detailed methodologies to calculate these figures are given.

Our main results are the following.

1) Despite the poor amount of information we require to build our MSI, it is significantly correlated to the other sentiment indices[5].

2) The correlation analysis shows that the survey-based French sentiment index is not significantly correlated to market returns. On the opposite, the two Baker-Wurgler indices and the MSI and BSI indices are strongly correlated to market returns, either when considering contemporaneous or lagged correlations. In other words, these indices can have a predictive power of future returns.

3) We use the two-step methodology of Baker-Wurgler (2006) to test the predictive power of sentiment indices. In a first step, these authors simply regress the returns of long-short portfolios based on size, on their sentiment index. In a second step, they control for the Fama-French-Carhart factors. We duplicate this approach on our data. We therefore perform regression analyses corresponding to the two above steps. The results show that the orthogonalized version of our MSI is always statistically significant, delivers the highest adjusted $\overline{R}^2$ in the uncontrolled and controlled regressions. It appears that the Buy-Sell imbalance index is outperformed by MSI in the two cases, despite the fact that MSI "neglects" trades which do not change diversification levels.

4) The performance of MSI is robust to variations in the minimum portfolio value of investors included in the database. Including constraints to keep only portfolios worth more than 1,000€ or 5,000€ only slightly reinforce the results. Our index is also robust to the arbitrary choice of the number of states of the Markov chain. Perhaps surprisingly, a lower maximum number of stocks reinforces the results. More precisely, when all investors with more than n stocks are seen as identical in terms of diversification, the predictive power of the MSI index is higher for $n = 10$ than for $n = 20$ or 30. It shows that the strength of the sentiment signal sent by changes in the diversification level of a portfolio

[5] Baker, Wurgler and Yuan (2009) show the contagion between indices in six countries, including France and the U.S. It is then not too surprising that French and U.S indices are positively correlated. The only exception is the correlation between MSI and the first Baker-Wurgler index which is positive but not significant at the 10 % level.

is larger when the portfolio is less diversified.

The remainder of the paper is organized as follows. Section 1 describes the diversification dynamics as a Markov chain and shows that trading conditions allow for the existence of a steady-state equilibrium. We then define our market sentiment index after having illustrated the intuition behind this approach. Section 2 presents the data and some descriptive statistics. Section 3 contains the main empirical results and section 4 provides several robustness checks. A final section concludes and proposes further directions of research.

1 Diversification level dynamics

1.1 The Markov chain of diversification levels

In this section we show how the dynamics of the number of stocks in the portfolio of individual investors can be described by a Markov chain. We assume that K stocks are traded on the market by I investors. The market is open at dates $t = 1, 2, ..., T$. As mentioned in the introduction, we know the composition of portfolios of individual investors at given points in time (on a monthly basis in this study). We are then able to evaluate the variation of the number of different stocks in their portfolios between dates t and $t+1$, for $t = 0$ to T ($t = 0$ is the beginning of 1999 and $t = T$ corresponds to December 2006).

Let N_t^i denote the number of different stocks held by investor $i \in \{1, ..., I\}$ at date t. N_t^i can be seen as a random variable taking values in the set $\{1, ..., K\}$.

Q_t^i stands for the one-period transition probability matrix of the stochastic process $(N_t^i, t = 0, ..., T)$. It is defined by:

$$\forall 1 \leq k \leq K, \forall 1 \leq m \leq K, Q_t^i(k,m) = P\left(N_t^i = m \left| N_{t-1}^i = k\right.\right) \tag{1}$$

$Q_t^i(k,m)$ is the probability that the portfolio of investor i contains m different stocks at date t, knowing she held k different stocks at date $t-1$. From now on, we assume that investors are homogeneous in the sense that $Q_t^i = Q_t$ for all $i \in \{1, ..., I\}$. All lines in Q_t sum to 1, by construction of the transition probability matrix of a finite Markov chain. For the empirical analysis to follow, we assume that K is not too large($K = 20$ in the typical case). State K will receive all portfolios with a number of different stocks greater or equal to K.

The structure of Q_t gives an idea of the dynamics of portfolio diversification between $t-1$ and t. It is important to notice that Q_t does not carry any specific information about trading volumes or about which stocks are traded. Roughly speaking, if the terms above the diagonal of Q_t are greater than those below the diagonal, we expect an increase in the mean number of stocks in portfolios over time. If the opposite is true, the portfolio of the investor should be more concentrated (containing a lower number of different stocks) in future periods. Our measure of diversification is then really simple.

Other measures of diversification have been used in the literature like the Herfindahl index of weights in the portfolio (Mitton and Vorkink, 2007) or the normalized portfolio variance, defined as the ratio of the portfolio variance and the mean variance of stocks in the portfolio (Goetzmann and Kumar, 2008). These alternatives could also be used to test our model but it would require to define (arbitrary?) ranges of diversification levels to categorize investors and build the transition probability matrix. Moreover, it would not really be consistent with our interpretation in terms of sentiment. In fact, we said before that buying new stocks reveals optimism, but a change in the normalized portfolio variance or in the Herfindahl index do not mean that investors change something in their portfolio. A variation in these measures may simply appear because stock prices move over time.

If Q_t signals optimism/pessimism of investors between dates $t-1$ and t, a natural question is to know what would be the portfolios in the long-run if investor sentiment were remaining stable over time, in other words if Q_t was unchanged at future dates. The answer to this question can be easily obtained under mild technical conditions, thanks to the properties of homogeneous Markov chains[6].

1.2 Steady-state equilibrium of diversification levels

Homogeneous Markov chains have a nice property: it is possible to find a steady-state equilibrium, that is a vector $\pi' = (\pi_1, ..., \pi_K)$ such that π_k is the proportion of investors holding k stocks in the long run. However, for the vector π to exist, the two following conditions must be satisfied[7].

Denote $Q_t^{(n)}$ the n-period transition matrix defined by $Q_t^{(n)}(k,m) = P(N_{t+n} = m \,|N_t = k)$ for $(k,m) \in \{1, ..., K\}^2$.

1. The Markov chain has to be irreducible, that is for each pair (k,m) there exists n such that $Q_t^{(n)}(k,m) > 0$. It is generally said that k and m communicate.

2. The chain has to be aperiodic. Denote $\mathcal{R}(k) = \{n \in \mathbb{N}^* \text{ such that } Q_t^n(k,k) > 0\}$ the set of return times of state k. The period of k, denoted $p(k)$, is the greatest common divisor of the numbers in $\mathcal{R}(k)$. The chain is said aperiodic if $p(k) = 1$.

Conditions (1) and (2) are satisfied in our case because individual investors can buy new stocks or sell stocks they hold without reglementary constraints.

Q_t being assumed identical for all investors, its elements can be estimated by:

$$Q_t(k,m) = \frac{\sum_{i=1}^{I} \mathbf{1}_{\left\{N_{t+1}^i = m\right\} \cap \left\{N_t^i = k\right\}}}{\sum_{i=1}^{I} \mathbf{1}_{\left\{N_t^i = k\right\}}} \tag{2}$$

where $\mathbf{1}_A$ is the indicator of the event A, valued 1 if A is true and 0 otherwise.

The specific properties of Markov chains allow to simply evaluate the two-period transition matrix.

[6] A Markov chain is said homogeneous when Q_t does not depend on t.
[7] See for example Roger (2010), chapter 1.

Denoting as before $Q_t^{(2)}(k,m) = P\left(N_{t+2} = m \,|N_t = k\right)$, the Chapman-Kolmogorov equations imply:

$$Q_t^{(2)} = Q_t * Q_t = Q_t^2 \tag{3}$$

More generally, the n-period transition probability matrix satisfies $Q_t^{(n)} = Q_t^n$. The steady-state equilibrium is given by any line of the limit matrix $\lim_{n\to+\infty} Q_t^n$ because all lines of this matrix are equal.

1.3 An illustration

As an illustration, we provide hereafter the successive powers of two different (5,5) transition matrices Q and Q_*. The first one on the left of table 1 characterizes "optimistic" or high-sentiment investors and the second one on the right more "pessimistic" investors.

We simply provide the powers 1, 4 and 8 of the two matrices to illustrate the convergence process. Q (Q_*) is said "optimistic" (pessimistic) because, roughly speaking, the probabilities of increasing (decreasing) the number of stocks are higher than probabilities of decreasing (increasing) this number. When looking directly at matrices Q and Q_*, it may seem difficult to detect which matrix leads to an increase (decrease) in diversification in the long run. However, remember that transition matrices are evaluated with equation (2). A rough indicator showing if diversification increases is the ratio of the number of investors above the diagonal divided by the total number of investors. Suppose that in our example, there are 10,000 investors in each of the five lines of each matrix. The numbers in Q show that 5,000 investors with one stock at date $t-1$ increase their number of stocks at date t. The second line gives 3,500 investors increasing diversification. Globally, 16,500 investors increase diversification between the two dates, that is 33% of the total number of investors in this sample. For Q_*, the same calculation gives only 10,000 investors (20%) increasing diversification.

[Insert Table 1 around here]

Even for $n = 8$, we observe that the equilibrium is not yet reached because slight variations still appear across lines. In fact, the true equilibrium distribution leads to 11.3% of investors with only one stock, 22.3% with 2 stocks, and so on. Concerning Q_*, the initial matrix is not so different but probabilities of decreasing the number of stocks are a little bit higher. The long-run consequence of these differences is that the steady-state equilibrium gives 18.1% of single-stock owners, 29.6% of investors with two stocks, and so on.

The corresponding equilibrium proportions (vectors π and π_*) are equal to:

$$\pi = (0.113; 0.223; 0.239; 0.223; 0.202) \tag{4}$$

$$\pi_* = (0.181; 0.297; 0.209; 0.188; 0.126) \tag{5}$$

The cumulative distribution functions of π and π_*, denoted F and F_*, are then given by:

$$F = (0.113; 0.336; 0.575; 0.798; 1) \tag{6}$$

$$F_* = (0.181; 0.478; 0.687; 0.855; 1) \tag{7}$$

These two functions F and F_* are plotted on figure 1. The bold (dashed) line represents the pessimistic(optimistic) distribution. The optimistic distribution F stochastically dominates the other one at the first-order since F is always below F_*. It means that transitions induced by Q lead to more diversified portfolios than those driven by Q_*. According to these curves, a measure of market optimism (pessimism) is the area above the cumulative distribution. We will keep this way of measuring the market sentiment index (hereafter MSI). To normalize the index between 0 and 1, we divide the area by the maximum number of stocks K minus 1. In this example, we get the following values:

$$\begin{aligned} MSI &= 1-\frac{1}{4}\left(\frac{0.113+0.336}{2}+\frac{0.336+0.575}{2}+\frac{0.575+0.798}{2}+\frac{0.798+1}{2}\right)=0.434 \\ MSI_{*} &= 1-\frac{1}{4}\left(\frac{0.181+0.478}{2}+\frac{0.478+0.687}{2}+\frac{0.687+0.855}{2}+\frac{0.855+1}{2}\right)=0.347 \end{aligned}$$

[**Insert Figure 1 around here**]

In the empirical analysis, each transition matrix Q_t estimated with diversification changes between $t-1$ and t allows to calculate an equilibrium distribution F_t, and consequently the market sentiment index MSI_t.

1.4 A formal definition of the market sentiment index (MSI)

We are now ready to define in a more formal way our sentiment index as the area above the cumulative distribution function of the equilibrium number of different stocks in the portfolio.

Definition 1 *For a transition matrix Q_t between $t-1$ and t, denote $N_{\infty,t}$ the random variable "number of different stocks" in the steady-state equilibrium. The investor sentiment index MSI_t is defined by:*

$$MSI_t = 1 - \frac{1}{K-1}\sum_{k=1}^{K-1} \frac{(P(N_{\infty,t} \leq k) + P(N_{\infty,t} \leq k+1))}{2} \tag{8}$$

As said before, MSI_t simply measures the area above the cumulative distribution function of $N_{\infty,t}$. It is important to remind that the essential feature of the convergence theorem of Markov chains is that the steady-state equilibrium does not depend on the initial distribution of investors.

In particular, it means that, Q_t being given, we do not need to know what is the sharing of the sample of investors between single-stock holders, two-stock holders, and so on, to evaluate the sentiment index. Only the changes between $t-1$ and t are important. Of course, it does not mean that Q_t is independent of the distribution of investors at $t-1$. Intuitively, the probability of decreasing diversification is different when investors hold respectively 3 or 30 stocks.

Q_t contains useful information. For example, it has been observed during long bullish high-sentiment periods (like during the dotcom bubble), that more and more investors enter the market and the ones already in increase their stakes, investing in new stocks, thus increasing diversification[8].

Roughly speaking, $Q_t\,(k, m) > Q_t(m, k)$ in bullish markets. In bearish markets or recession periods, investors are reluctant to put new money on the table and possibly sell stocks to finance consumption or liquidity needs. Consequently, we expect $Q_t\,(k, m) \leq Q_t(m, k)$ in bearish markets. However, some asymmetry may arise; in fact, at the individual level, a decrease in diversification does not always reveal pessimism. It is well known that individual investors are prone to the disposition effect, selling winners too early and riding losers too long[9]. Consequently, it may happen that bearish markets induce some inertia in the transition matrix, investors keeping their losing stocks. It turns out that the time-series of the terms on the diagonal of Q is a good indicator of pessimism.

As we focus on MSI as an indicator of sentiment, we need to take into account the potential relationship between MSI and other usual risk factors to assess the marginal contribution of sentiment in the explanation of returns. We then build an orthogonalized version of the MSI by taking the residuals of the regression of sentiment on Fama-French factors and Carhart momentum factor:

$$MSI_t = \alpha_0 + \alpha_{Mkt} RMRF_t + \alpha_S SMB_t + \alpha_H HML_t + \alpha_M MOM_t + \varepsilon_t \tag{9}$$

This orthogonalized MSI is denoted $MSI^{\perp}$.

[8] Goetzmann and Kumar (2008) point out an increase in the average number of stocks held by a large sample of U.S investors over the period 1991-1996 (which was almost always bullish). From 4.28 in 1991, it rises to 6.51 in 1996 but the authors do not attribute this variation to an increase in financial skills of retail investors.

[9] The disposition effect is one of the well-documented biases of individual investors. It has been first studied by Shefrin and Statman (1985). A number of empirical studies in several countries show that individual investors are prone to the disposition effect (Odean(1998) for the U.S, Shapira and Venezia (2001) for Israel, Barber *et al.*(2007) for Taiwan, Boolell-Gunesh *et al.* (2009) for France).

Notations are standard. $RMRF_t$ is the French market index return (in excess of the risk-free rate). It is given by the Eurofidai value-weighted general index (calculated with the methodology of the Center for Research in Security Prices (CRSP)). This index is based on around 700 stocks over the period under consideration. HML_t is the book-to-market factor and MOM_t the momentum factor. They are provided by Eurofidai and calculated according to the Fama-French (1993) methodology for HML_t and to the methodology of Carhart (1997) for the momentum factor MOM_t .

1.5 Buy-Sell imbalance

For a given stock i, the buy-sell imbalance index in month t is defined by:

$$BSI_{it} = \frac{\sum_1^{d_t} (B_{it} - S_{it})}{\sum_1^{d_t} (B_{it} + S_{it})} \tag{10}$$

where d_t is the number of trading days in month t, $B_{it}(S_{it})$ is the volume of buying (selling) trades for stock i in month t.

To define a portfolio BSI, we can either average the BSI of the stocks in the portfolio (see for example Kumar and Lee, 2006) or aggregate buy and sell trades on stocks in the portfolio. In this paper, following Barber and Odean (2008), we define the market BSI index by:

$$BSI_{Mt} = \frac{\sum_{i=1}^{N} \sum_1^{d_t} (B_{it} - S_{it})}{\sum_{i=1}^{N} \sum_1^{d_t} (B_{it} + S_{it})} \tag{11}$$

where $B_{it}(S_{it})$ is defined as before and N is the number of different stocks traded in month t.

This definition differs from the one given by Kumar and Lee (2006) since they calculate an average BSI of individual stocks. We prefer the definition given in equation 11 since we focus on a market sentiment index, not on stock-level sentiment indices. Given the trade frequency of investors in the database, 25 % to 30 % of stocks are traded each month; consequently stock $BSIs$ cannot be defined for every stock in every month. It justifies the way we calculate the aggregate buy-sell imbalance index. As usual, when comparing means of ratios and ratios of means, the difference between the two definitions is increasing in the cross-sectional variation of stock-level $BSIs$.

Moreover, to be consistent in our comparison with the other indices, especially MSI, we cumulate in a month t only the trades of investors who are still in the database the day after they trade. In fact, when investors leave, we cannot know why. Maybe they simply change their broker or they need money to buy a house or their preceding losses convince them to give up investment in stocks.

As for MSI, we also consider an orthogonalized bersion of BSI by taking the residuals of the following regression

$$BSI_t = \beta_0 + \beta_{Mkt} RMRF_t + \beta_S SMB_t + \beta_H HML_t + \beta_M MOM_t + \varepsilon_t \tag{12}$$

The orthogonalized index is denoted $BSI^{\perp}$ and will also be considered in the analyses to follow.

The definition of BSI takes all trades into account to measure sentiment. But as shown by Statman et al. (2006), trading volume is partly driven by overconfidence and the disposition effect, themselves influenced by past market returns. It turns out that an increase of buy trades can reveal a reinforcement of optimism when an increase in sales is due to the disposition effect when past returns are positive. It is then worth to notice that an equal increase in buy and sell trades decrease the BSI value. In fact, the numerator is unchanged while the denominator is increased by the amount of trade on both sides. It is then unclear whether the definition of BSI is able to measure euphoria or high sentiment during bullish markets.

2 Data and descriptive statistics

2.1 Investors data

Data on individual investors come from a large French brokerage house. We obtained transaction data for all active accounts over the period 1999-2006, that is a total of around nine million trades, for 92,603 investors. The trades file combines the following information for each trade: ISIN code of the asset, buy-sell indicator, date, quantity and amount in Euros. In the investors file, some demographical characteristics of investors are gathered: date of birth, gender, date of entry in and exit of the database, opening and/or closing dates of all accounts and region of living.

Some investors open an account within this period, some others close their account before the end of the period. As it would make no sense to analyze portfolios every day (due to the low turnover of portfolios), we chose to "take a photograph" of portfolios at the end of each month. It turns out that some investors may hold no position in a given month, even if they held a portfolio before, and restart to trade after. We deleted investors with positions on stocks for which price data were not available for at least one year and portfolios worth less than 100 €. Finally, 87,373 investors were considered in the analysis (they held stocks at least two successive months in the period) but their number varies over time. 8,258,809 trades remain in our final database. On average, the number of investors in a month is 51,340 with a minimum of 34,230 and a maximum of 60,001.

Figure 2 shows three time-series. The upper dotted curve represents the average number of stocks held by investors ($\times 10^4$). We observe that underdiversification is more the rule than the exception since the average number of stocks varies from 5.5 to 6.8, and the median is 3 or 4 all over the period. The difference between median and mean is explained by a low percentage of investors holding largely diversified portfolios. These figures are in line with the ones obtained in the studies referred to in the introduction. It is then reasonable to postulate that individual investors hold underdiversified portfolios and that buying a fourth stock when three are already in the portfolio has not the same meaning as buying a new stock when the portfolio is already fully diversified with 200 stocks. The striking feature of this curve is the sharp increase of the average number of stocks just before the dotcom bubble burst, that is during the first months of 2000. Then a decrease to the former level of diversification is observed. On the remainder of the period, the average number of stocks is roughly stable. This curve also shows that considering the average change of diversification would not bring enough information. It is the reason why we use the Markov chain technology to better extract information about sentiment. The evolution of the number of stocks is different from the one observed by Goetzman and Kumar (2008) on a sample of U.S investors. As mentioned before, they found an increase in diversification over the period 1991-1996 because the market was bullish almost all the time.

The middle bold curve and the bottom dashed curves provide the evolution of the mean and median portfolio values. The first month being January 1999, it appears that the average portfolio value follows the evolution of the market as a whole. A sharp increase in value appears in the 15 first months, up to the Internet bubble burst in April 2000. Then, portfolio values decrease until April 2003 (the market bottom), and finally a partial recovery is observed between 2003 and the end of our period (December 2006). Consequently, the evolution of portfolio values in our sample does not seem different from the evolution of the stock market.

[**Insert Figure 2 around here**]

As for the number of stocks in portfolios, there is a large discrepancy between the mean and median portfolio values, a result in line with other studies on individual investors (for example Mitton and Vorkink, 2007). In fact, a few investors are very wealthy, compared to the average investor; they move upward the average portfolio value in a significant way. On average, 0.2% of investors hold a stock portfolio worth more than one million euros.

Table 2 gives some more detailed statistics at three points in time, January 2000, January 2003 and January 2006[10]. We use the same presentation as Table 2 of Mitton-Vorkink (2007). At the end of

[10]The complete statistics for all months of the period are available upon request.

each month, we divide investors into seven categories (first column of Table 2) . The first five contain investors holding one to five stocks, the sixth groups investors with six to nine stocks and the last category groups all diversified investors with ten stocks or more. The second column gives the number of investors in each category. The four last columns provide summary statistics about portfolio values, namely the mean, the first quartile, the median and the third quartile. There is a large proportion (around 20%) of single-stock owners and in all categories, the mean portfolio value is much higher than the median, even among single-stock owners. It reinforces the preceding remark about figure **2.** In most cases, the mean is close to the third quartile. These observations are similar to the results of Mitton and Vorkink (2007) on a large sample of U.S investors.

[**Insert Table 2 around here**]

The market activity of investors in our sample is also highly variable over time. Figure 3 shows the time-series of monthly trades. The bold (dashed) line represents buy(sell) trades. The large variations are essentially observed in the three first years with a dramatic increase in the two kinds of trades up to April 2000. Around 110,000 monthly buy trades were realized in February, March and April 2000. An equivalent decrease is then observed until September 2001. Of course, even if the French market remained open after the 9/11, the volume was considerably lower that month. The remarkable fact is that sales were also at a very low level. It is not clear what a BSI index means in this special case. In fact, it is valued 0.03, that is largely above the median which is 0.0046. On the contrary, the same figures for MSI are 0.0694 for September 2001 with a median equal to 0.189.

In the last five years of our sample period, the average level of trades is around 35,000 trades a month on each side.

[**Insert Figure 3 around here**]

2.2 Stock data

Stock prices come from two sources, Eurofidai for stocks traded on Euronext and Bloomberg for the other stocks. We used daily prices for estimating the moments of the distribution of returns on stocks and investors portfolios. In our sample, the universe of investments contains 2,491 stocks, meaning that each of these stocks has been traded at least once over the period. There are 1,191 French stocks, the remaining coming from all over the world but essentially from the U.S (1,020 stocks), United Kingdom (62), Netherlands (34), Germany (31) and Italy (15). Despite the large number of U.S stocks in our sample, the trades on French stocks count for more than 90% of the trading volume, as shown on panel A of table 3. It illustrates the well-known home bias puzzle[11]. It is the reason why most comparisons in this paper are related to the French market. Moreover, if we compare the number of U.S stocks to the volume of trade on these stocks, we observe that they are very unfrequently traded. Only 54,881 trades on U.S stocks were executed, compared for example to the 366,138 trades on the 34 Dutch stocks. Concerning holdings, panel B of table 3 reports at the end of each year from 1999 to 2006 the proportion of investors holding stocks of the 6 main countries in the database. For example, at the end of 2003, there were 56,952 investors holding stocks. 96.97% held French stocks (meaning that around 3% held only foreign stocks), 21.05% held Dutch stocks but only 3.97% held U.S stocks, despite the large number of U.S stocks in the database (held at least once during the period).

[Insert Table 3 around here]

3 Empirical study

3.1 Correlation analysis

In this section we compare the MSI index to four other indices, namely, the French sentiment index (FSI), the two indices developed by Baker and Wurgler (2006), denoted $BW1$ and $BW2$, and the

[11] See Lewis (1999) and Karolyi and Stulz (2003) for a literature review on this topic.

Buy-Sell imbalance index (BSI). The correlations are given for the two versions (orthogonalized or not) of MSI and BSI. We also analyze the contemporaneous correlation between these indices and the market index.

The correlations are reported in table 4. MSI and BSI are computed on a monthly basis using our sample of investors. MSI_t is obtained with a one-month transition matrix Q_t based on the sub-sample of investors present in the database at dates $t-1$ and t. BSI for month t is calculated by cumulating daily trades over the month. However, for a trade on day s to be considered, the investor must still be in the database at the end of the day. As mentioned when we defined BSI, trades of people leaving the database on a given day are not taken into account.

All indices are highly positively correlated, the only exception being the correlation between MSI and the first Baker-Wurgler index which is not significant even if it is positive. These positive correlations are not a surprise, even when comparing French and U.S indices. It was already mentioned in Baker, Wurgler and Yuan (2009). They showed that there is some contagion between indices in six countries, including France and the U.S. They found a .44 correlation between global sentiment indices of the two countries. The fact that MSI is significantly correlated to the other indices is the first remarkable result according to the "poor" information used to calculate our index (at least according to the standards of classical finance theory). No information is used about which stocks are traded, nothing is known neither about prices at which people trade nor about trading volumes. Moreover, due to the convergence theorem of Markov chains, MSI_t does not depend on the diversification levels at date $t-1$ since it is calculated with the variations of diversification between $t-1$ and t.

The question is then to know if there is something different in MSI compared to the other usual indices. The answer appears in the two last columns of table 4. MSI is the only index with a positve contemporaneous correlation with the market return. Even if we compare with BSI (which is calculated with the same basic data, that is trades of individual investors), we observe a large positive correlation between the two (0.747) but BSI is negatively correlated (but not significantly) to the market return (-0.104) when MSI is significantly positively correlated with $RMRF$ (0.249). Moreover, lagged correlations between MSI, $MSI^{\perp}$ and the market return are still significant and positive. It is no more the case for the BSI index. The most interesting point is that the orthogonalized version of MSI is significantly correlated to future returns, allowing to enter a multifactor approach to study the predictive power of our index and to perform a systematic comparison with the other indices.

[**Insert Table 4 around here**]

3.2 The multi-factor approach

3.2.1 Predicting returns on size portfolios

In this section, we compare the market sentiment index MSI to the others through predictive regressions on size portfolios. Baker and Wurgler (2007) introduced the "sentiment seesaw" to explain the effect of sentiment on stocks (figure 1, p133). They showed that sentiment can have opposite effects on stock returns, depending on the difficulty to arbitrage. In high-sentiment periods, large stocks may be undervalued and small stocks overvalued. The reverse appears in low sentiment periods. If this theoretical prediction is true, a good sentiment measure should help to predict returns. We then consider a long-short portfolio with a long position on small caps (more difficult to arbitrage) and a short position on big caps (easier to arbitrage). According to the "sentiment seesaw" approach, we expect the portfolio return to be high following low sentiment periods and to be low following high-sentiment periods. In other words, when regressing the return of the long-short portfolio, we expect a negative sign for the coefficient of the sentiment measure.

Following Baker and Wurgler (2006), we consider two steps. The first one corresponds to the following equation:

$$R_{Smallcaps,t} - R_{Bigcaps,t} = a + bSENTIMENT_{t-1} + \varepsilon_t \tag{13}$$

where $SENTIMENT_t$ is the sentiment index for month t and may be FSI, $BW1$,$BW2$, MSI,BSI,$MSI^{\perp}$ or $BSI^{\perp}$.

In the second step, we control for Fama-French and Carhart factors (except the size factor since the long-short portfolio is based on size). The regression model is then the following:

$$R_{Smallcaps,t} - R_{Bigcaps,t} = c + dSENTIMENT_{t-1} + \beta RMRF_t + hHML_t + mMOM_t + \varepsilon_t \qquad (14)$$

[**Insert Table 5 around here**]

The results appear in table 5 with Newey-West consistent estimates. Panel A (B) provides the regression coefficients of the sentiment measures for equation 13 (14) without (with) control for the Fama-French-Carhart factors. In all cases, we get the expected sign for the regression coefficient, that is negative. It means that, on average, a period of high sentiment is followed by a low return on the long short portfolio, even after controlling for the market return, book-to-market and momentum factors. In the two versions of the analysis, the orthogonalized MSI is significant and delivers the highest $\overline{R}^2$ even if the significance level of Baker-Wurgler indices is higher in this case.

These results show that portfolio diversification dynamics of individual investors carry valuable information to predict returns. As mentioned in the introduction, this measure can be easily implemented and updated regularly by banks using their own portfolio of retail clients.

3.2.2 Are BW indices and MSI substitutes or complements?

In section 3.1, we observed that BW indices are negatively correlated to returns when MSI is positively correlated to the same returns. At the same time, the correlation between BW and MSI, though positive, is low. It suggests that the two indices possibly measure different dimensions of sentiment. If it is the case, introducing the two in the regression analysis of the former section could improve the results. Table 5 showed that $BW1$ and $BW2$ are significant in predicting returns of the long-short portfolio when the Fama-French and Carhart factors are introduced as control variables. When we introduce the two variables $BW1$ and $MSI^{\perp}$ in this regression, the two coefficients are significant at the 5 % level and the adjusted R^2 is equal to 0.261. When only one of the two sentiment measures was considered this statistic was 0.225 for $BW1$ and 0.246 for $MSI^{\perp}$ (see Panel B in table 5) . This is a slight improvement but it comes at a cost: increasing the number of sentiment measures to consider.

4 Robustness checks

In this section, we analyze the sensitivity of MSI and $MSI^{\perp}$ to the choices made to build these measures. More precisely, up to now indices were obtained by assuming a Markov chain with 20 states and no constraint was imposed on the portfolio value of investors, that is we kept in the database all investors with portfolios valued more than 100€. In the following we test if different choices for these variables deteriorate or improve the results.

4.1 The number of states of the Markov chain

In the preceding section, we arbitrarily chose $K = 20$, meaning that all portfolios containing more than 20 stocks at two successive dates are considered as unchanged in terms of diversification. It is then worth to check is this choice is important in getting the results. Table 6 is built as table 5 except that we only keep indices MSI and $MSI^{\perp}$ and vary the number K from 10 to 30 by steps of 10. The results concerning $K = 20$ are recalled for the ease of reading. The significance of regression coefficients is not strongly influenced by the maximum number of stocks K. However, it is worth to notice that the significance of the coefficient of $MSI^{\perp}$ is reinforced when $K = 10$. A possible interpretation for this result is that a change in diversification levels has a stronger meaning in terms of sentiment for portfolios with a low number of stocks. If buying a new stock reveals optimism, it is probably more striking when an investor's portfolio goes from 2 to 3 stocks than from 18 to 19. Moreover, the way we calculate the index gives less weight to changes in diversification of portfolios already containing a "large" (but lower than K) number of stocks.

Of course, there is a limit to this inverse relationship between the number of states of the Markov chain and the significance of regression coefficients. More precisely, the case $K = 5$ (not reported) shows a deterioration of the estimation. It is not surprising since the information contained in the transition matrix of the Markov chain is really insufficient when K is too low.

[**Insert Table 6 around here**]

4.2 The minimum portfolio value

A second robustness check needs to be performed since we imposed almost no constraint on the portfolio value. We kept all investors with a portfolio worth more than 100 €. The fact that investors holding such portfolios are included in the calculation of MSI raises the question of knowing if they really are "sentiment traders". Liu (2010) develops a model showing that underdiversification of individual portfolios is essentially justified by wealth levels and solvency constraints. In short, "poor" investors subject to solvency constraints only focus on expected returns and not on variance of returns because their wealth needs to be invested for a large part in the risk-free asset. Consequently, the optimal risk-return tradeoff for them is to invest the remainder of their wealth in high expected return stocks, also meaning highly risky stocks.

Typically, an investor with only a few hundred euros to invest in the stock market will buy a single stock. When wealth increases, a second stock is introduced in the portfolio because the marginal benefit of underdiversification is compensated by the supplement of risk generated by the single-stock portfolio.

This kind of reasoning has nothing to do with sentiment but relies on simple and maybe convincing arguments. Consequently, to check if our sentiment measure really measures sentiment and not solvency constraints, we restrict our database to investors whose portfolio value is at least 1,000€ and 5,000€. We recalculate MSI and $MSI^{\perp}$ in each case and perform one more time the preceding analyses.

The results appear in tables 7. Increasing the minimum portfolio value has essentially two effects. It decreases the number of investors included in the analysis and increases the mean number of stocks in portfolios. For example, the average number of investors becomes 42,355 (27,878) when the minimum portfolio value is 1,000€ (5,000€), instead of 51,340 without constraints. These figures remain sufficient to evaluate transition matrices with a good accuracy but it changes the long-run equilibrium distribution of the Markov chain of diversification levels, and consequently the sentiment measures MSI, $MSI^{\perp}$. The mean numbers of stocks in portfolios are respectively equal to 5.92, 6.8 and 8.92 for the minimum portfolio values of 100€, 1,000€ and 5,000€. It gives some credibility to Liu's assumption and makes necessary the sensitivity analysis performed here. Moreover, looking at single-stock owners shows that they are on average 11,192 in the complete sample but only 5,092 (1,343) when portfolio value is larger than 1,000€ (5,000€). In table 7, the coefficients of MSI increase when the minimum portfolio value increases but such a variation is not present for $MSI^{\perp}$. The statistical significance of the coefficients is roughly the same in the three situations for $MSI^{\perp}$. Moreover, there is no clear influence of this minimum portfolio value constraint on the adjusted $\overline{R}^2$ of the regression. In the controlled regression, $\overline{R}^2$ slightly increases with the portfolio value but it slightly decreases in the "uncontrolled" regression. As a conclusion, we can say that even if the average number of stocks increases when minimum portfolio values are higher, there is no significant change in the diversification dynamics. It confirms the robustness of our market sentiment index, at least in the orthogonalized version.

[**Insert Table 7 around here**]

5 Concluding remarks

This paper proposes an original measure of market sentiment based on changes in diversification choices of retail investors. The purpose is to extract information about sentiment (optimism/pessimism) of retail investors and to show that this sentiment measure enters significantly in the short-term prediction of market returns. We show that this new measure outperforms several other indices (based on surveys or on macroeconomic variables) in this task. Our contributions are theoretical and empirical. To the best of our knowledge, the sentiment index we introduce in this paper is completely new and we show that the Markov chain technology allows to better extract information about sentiment from the data. On the empirical side, our index can be easily implemented by banks using the accounts of their own clients. It allows to follow dynamically the market sentiment and to update it frequently.

However, in this paper we did not address a complex issue linked to the disposition effect. More precisely, a decrease in diversification does not always signal a worsening in sentiment. It can be due to the realization of gains by investors prone to the disposition effect. In fact, the proportion of investors decreasing diversification is positively correlated (even if this correlation is not significant) with the proportion of investors increasing diversification. In other words, the proportion of investors who do not change diversification is a better indicator of pessimism. This problem also affects sentiment indices based on buy-sell imbalances, possibly in a stronger way because such indices are calculated as a difference between purchases and sales normalized by the sum of the two. Consequently, investors doing nothing do not enter the picture. But the reluctance to sell losers may lead to consider those who do not trade as pessimistic investors. In short, more work is needed on this point.

Finally, the empirical study proposed in this paper concerns a large sample of French individual investors. To confirm the interest of our index, it would be useful to duplicate and enrich the test on other samples of investors, especially in other countries.

References

[1] Andrade, S., Chang, C., Seasholes, M. S. (2008): Trading imbalances, predictable reversals, and cross-stock price pressure. *Journal of Financial Economics*, 88(2), 406-423.

[2] Baker, M., Wurgler, J. (2006): Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, 61(4), 1645-1680.

[3] Baker, M., Wurgler, J. (2007): Investor Sentiment and the Stock Market. *The Journal of Economic Perspectives*, 21(2), 129-151.

[4] Baker, M., Wurgler, J. and Yuan, Y. (2010), Global, Local, and Contagious Investor Sentiment, Working Paper, Harvard Business School.

[5] Barber, B. M., Odean, T. (2008), All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors, *Review of Financial Studies*, 21(2), 785-818.

[6] Barber, B. M., Lee, Y., Liu, Y., Odean, T. (2009): Just How Much Do Individual Investors Lose by Trading? *Review of Financial Studies*, 22(2), 609-632.

[7] Barber, B. M., Lee, Y., Liu, Y., Odean, T. (2007): Is the Aggregate Investor Reluctant to Realise Losses? Evidence from Taiwan, *European Financial management,* 13(3), 423-447.

[8] Barber, B. M., Odean, T., Zhu, N. (2009): Do Retail Trades Move Markets? *Review of Financial Studies*, 22(1), 151-186.

[9] Blume, M. E., Friend, I. (1975), The Asset Structure of Individual Portfolios and some Implications for Utility Functions, *Journal of Finance*, 30(2), 585-603.

[10] Boolell-Gunesh, S., Broihanne, M.H., Merli, M. (2009), Disposition Effect, Investor Sophistication and Taxes: Some French Specificities, *Finance*, 30(1), 51-78.

[11] Broihanne, M.H., Merli, M., Roger, P. (2011), Underdiversification of French Individual Investors' Portfolios and Skewness Seeking, International Conference of the French Finance Association.

[12] Carhart, M. M. (1997): On Persistence in Mutual Fund Performance, *The Journal of Finance*, 52(1), 57-82.

[13] Calvet, L.E., Campbell, J.Y., Sodini, P. (2007), Down or Out: Assessing the Welfare Costs of Household Investment Mistakes, *Journal of Political Economy*, 115, 707-747.

[14] Dorn, D., Huberman, G., Sengmueller, P. (2008): Correlated Trading and Returns. *The Journal of Finance*, 63(2), 885-920.

[15] Fama, E.F., French, K. (1992): The cross-section of expected stock returns. The Journal of Finance, 47(2), 427-465.

[16] Foucault, T., Sraer D., Thesmar, D. (2011), Individual Investors and Volatility, *Journal of Finance*, 66(4), 1369-1406.

[17] Goetzmann, W.N., Kumar, A. (2008), Equity Portfolio Diversification, *Review of Finance*, 12, 433–463.

[18] Hvidkjaer, S. (2008): Small Trades and the Cross-Section of Stock Returns. *Review of Financial Studies*, 21(3), 1123-1151.

[19] Kaniel, R., Saar, G., Titman, S. (2008): Individual Investor Trading and Stock Returns.*The Journal of Finance*, 63(1), 273-310.

[20] Karolyi, A., Stulz, R.M. (2003), Are Financial Assets Priced Locally or Globally?, Handbook of the Economics of Finance, G. Constantinides, M. Harris, and R.M. Stulz, eds. Elsevier North Holland.

[21] Kelly, M. (1995), All their Eggs in one Basket: Portfolio Diversification of U,S Households, *Journal of Economic Behavior and Organization*, 27(1), 87-96.

[22] Kumar, A. and Lee, C.M.C. (2006), Retail Investor Sentiment and Return Comovements, *Journal of Finance*, 61(5), 2451-2486.

[23] Kumar, A. (2007), Do the Diversification Choices of Individual Investors Influence Stock Returns? *Journal of Financial Markets*, 10, 362-390,

[24] Lease, R,C., Lewellen, W.G., Schlarbaum, G.G. (1974), The individual investor: attributes and attitudes, *Journal of Business*, 29, 413–433.

[25] Lewis, K.K. (1999), Trying to Explain Home Bias in Equities and Consumption, *Journal of Economic Literature*, 37(2), 571-608.

[26] Liu, H, (2010), Solvency Constraint, Underdiversification, and Idiosyncratic Risks, *Working Paper*, Washington University in St, Louis.

[27] Mitton, T., Vorkink, K. (2007), Equilibrium Underdiversification and the Preference for Skewness, *Review of Financial Studies*, 20, 1255–1288.

[28] Odean, T. (1998): Are Investors Reluctant to Realize Their Losses? *The Journal of Finance*, 53(5), 1775-1798.

[29] Odean, T., (1999), Do Investors Trade Too Much?, *American Economic Review*, 89, 1279–1298.

[30] Pan, J., Poteshman, A. M. (2006): The Information in Option Volume for Future Stock Prices. *Review of Financial Studies*, 19(3), 871-908.

[31] Roger, P. (2010), *Stochastic Processes for Finance*, Ventus Publishing Aps, Frederiksberg, Denmark.

[32] Shapira, Z., Venezia I. (2001). Patterns of Behavior of Professionally Managed and Independent Investors. *Journal of Banking and Finance,* 25, 1573-1587.

[33] Schmitz, P., Glaser, M., Weber, M. (2007). Individual Investor Sentiment and Stock Returns - What Do We Learn from Warrant Traders? Working Paper, Universität Mannheim.

[34] Shefrin, H., Statman, M. (1985). The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. *The Journal of Finance*, 40(3), 777-790.

[35] Simkowitz, M. A., Beedles, W, L, (1978) Diversification in a Three-Moment World, *Journal of Financial and Quantitative Analysis* 13, 927–941.

[36] Statman, M., Thorley, S., Vorkink, K., (2006) Investor Overconfidence and Trading Volume, *Review of Financial Studies*, 19(4), 1531-1565.

[37] Weber, M., Camerer C. F. 1998. The Disposition Effect in Securities Trading: An Experimental Analysis. *Journal of Economic Behaviour and Organization,* 33, 167-184.

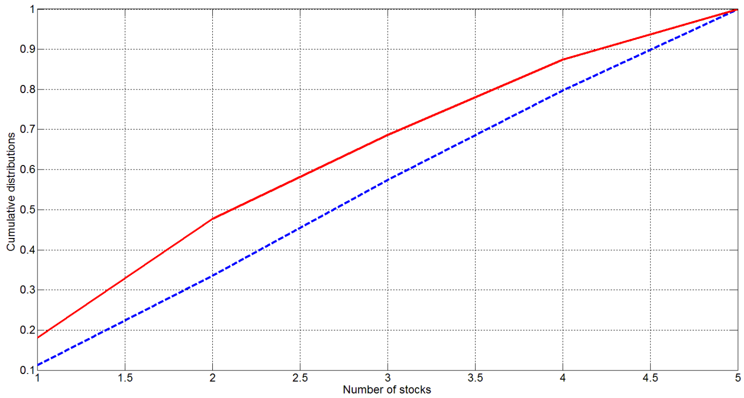

Figure 1
Cumulative long-run distributions F and F_* corresponding to transition matrices Q (dashed line) and Q_* (bold line).

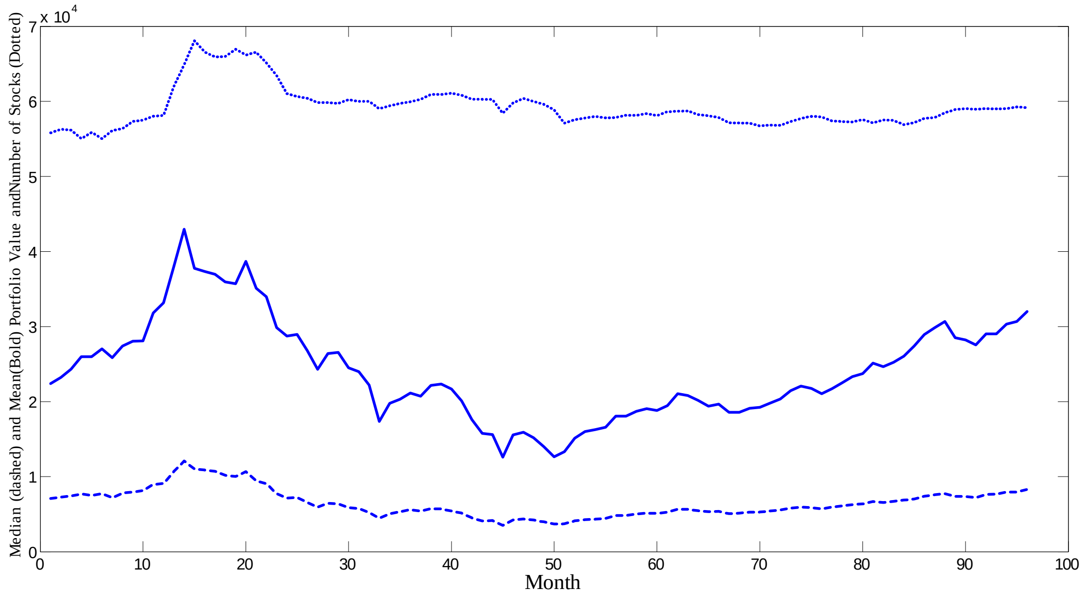

Figure 2
The three curves represent respectively the time-series of the average number of stocks held by investors, and the mean and median portfolio value. The period under consideration starts in January 1999 (month 1) and ends in December 2006 (month 96). The upper dotted curve is the average number of stocks ($\times 10^4$). The middle bold curve is the average portfolio value and the lower curve is the median portfolio value.

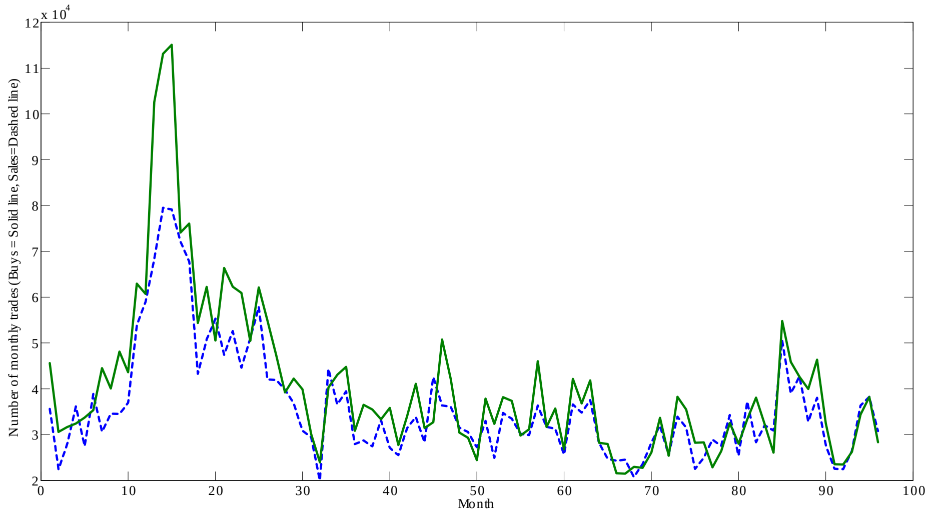

Figure 3
Time-series of the number of monthly trades. The solid (dashed) line represents the evolution of purchases (sales)

Power	Number	Q^n					Q_*^n				
	of stocks	1	2	3	4	5	1	2	3	4	5
$n=1$	1	0.5	0.2	0.1	0.1	0.1	0.5	0.2	0.1	0.1	0.1
	2	0.05	0.6	0.2	0.1	0.05	0.15	0.6	0.1	0.1	0.05
	3	0.1	0.1	0.5	0.2	0.1	0.1	0.2	0.5	0.1	0.1
	4	0.05	0.1	0.15	0.5	0.2	0.1	0.15	0.2	0.5	0.05
	5	0.05	0.1	0.15	0.15	0.55	0.05	0.1	0.15	0.2	0.5
$n=4$	1	0.143	0.245	0.226	0.204	0.182	0.199	0.293	0.197	0.182	0.129
	2	0.107	0.265	0.249	0.210	0.168	0.189	0.324	0.194	0.177	0.116
	3	0.117	0.208	0.247	0.230	0.198	0.174	0.293	0.223	0.182	0.129
	4	0.105	0.202	0.233	0.239	0.221	0.174	0.285	0.220	0.203	0.118
	5	0.105	0.202	0.233	0.224	0.236	0.156	0.262	0.219	0.208	0.155
$n=8$	1	0.114	0.225	0.239	0.222	0.200	0.181	0.296	0.208	0.188	0.126
	2	0.113	0.225	0.240	0.223	0.199	0.182	0.298	0.208	0.187	0.126
	3	0.113	0.222	0.239	0.224	0.203	0.181	0.296	0.209	0.188	0.126
	4	0.112	0.221	0.239	0.224	0.204	0.181	0.296	0.209	0.188	0.126
	5	0.112	0.221	0.239	0.224	0.204	0.179	0.295	0.210	0.189	0.127

Table 1
This table gives the powers 1, 4 and 8 of the (5,5) transition matrices Q ("more optimistic") and Q_* ("less optimistic"). The second column (line) gives the number of stocks at date $t(t+n)$.

Number of Stocks	Number of Investors	Mean Port. Value	1st quartile	Median Port. Value	3rd quartile
Panel A: Portfolios as of January 2000					
1	8749	7384, 88	766, 60	1751, 20	4332, 67
2	6635	10797, 04	2096, 11	4020, 80	8325, 20
3	5355	15392, 53	3703, 88	6415, 48	12703, 24
4	4175	21854, 06	5556, 14	9542, 00	18016, 15
5	3174	26407, 74	7306, 75	12482, 10	22922, 00
6 to 9	8378	42463, 71	11576, 10	19518, 49	36201, 08
More than 10	8747	110538, 37	28414, 02	50784, 95	95227, 55
All	45213	37961, 97	3613, 18	10693, 36	29753, 23
Panel B: Portfolios as of January 2003					
1	11379	2186, 86	234, 00	525, 20	1378, 80
2	7849	3730, 13	716, 87	1456, 40	3130, 47
3	6062	6109, 57	1323, 40	2514, 27	5221, 07
4	4803	7414, 78	2030, 81	3748, 70	7458, 02
5	3681	10692, 89	2951, 17	5179, 59	9926, 23
6 to 9	9144	15841, 77	4898, 52	8541, 48	15948, 40
More than 10	9793	43600, 66	13053, 03	23842, 09	44787, 36
All	52711	14001, 07	1173, 93	3975, 36	12224, 47
Panel B: Portfolios as of January 2006					
1	11229	4197, 63	399, 67	1039, 13	2651, 36
2	7393	8265, 72	1271, 75	2843, 28	6406, 62
3	5424	11287, 21	2470, 55	4984, 51	10784, 48
4	4126	16965, 01	3882, 90	7414, 97	15169, 86
5	3245	22977, 15	5352, 08	9891, 80	20385, 77
6 to 9	7983	32900, 79	9158, 55	16819, 18	32866, 52
More than 10	8089	89583, 57	24549, 25	45979, 47	92192, 24
All	47489	27402, 40	1957, 91	7017, 93	22815, 86

Table 2
Statistics on portfolio values at three points in time, January 2000 (Panel A), January 2003 (Panel B) and January 2006 (Panel C). The first column gives the way portfolios are categorized with respect to the number of stocks. Portfolios containing 6 to 9 stocks are in the same category and portfolios with more than ten stocks are also grouped. The second column shows the number of investors in each diversification group. The four last columns describe portfolio values by providing the mean portfolio value, the first quartile, the median and the third quartile

Panel A: Trades in stocks of the 6 main countries							
	Total	France	Netherlands	U.S.A	Great Britain	Germany	Italy
Number of stocks	2,491	1,191	34	1,020	62	31	15
Number of trades	8,258,809	7,510,017	366,138	54,881	27,207	22,849	5,059
Panel B: Percentage of investors holding stocks of the 6 main countries							
End of year	**Ninvestors**	**FR**	**NL**	**US**	**GB**	**DE**	**IT**
1999	43,638	98.32	6.10	4.50	1.48	2.64	0.27
2000	58,699	96.93	23.37	3.90	3.04	2.05	0.23
2001	57,587	97.16	21.74	3.61	1.52	2.02	1.29
2002	53,040	97.06	21.33	3.85	1.64	1.85	0.61
2003	56,952	96.97	21.05	3.97	1.61	1.19	0.70
2004	52,050	97.17	20.21	3.89	1.72	1.17	0.41
2005	47,937	97.82	13.82	3.30	1.80	1.19	0.08
2006	42,100	98.13	14.69	2.75	2.18	0.98	0.14

Table 3
Trades and holdings for stocks of the 6 main countries. The first line gives the number of stocks held at least once by investors in the sample and the second line the number of trades. The remainder of the table gives the percentage of investors holding stocks of the six countries at the end of each year of the period 1999-2006.

	$BW1$	$BW2$	MSI	BSI	$MSI^{\perp}$	$BSI^{\perp}$	$RMRF_t$	$RMRF_{t+1}$
FSI	0.742***	0.767***	0.421***	0.361***	0.444***	0.374***	-0.082	-0.130
$BW1$		0.957***	0,145	0.227**	0.204**	0.223**	-0.207**	-0.174*
$BW2$			0.185*	0.281***	0.240**	0.260**	-0.231**	-0.179*
MSI				0.747***	0.943***	0.756***	0.249**	0.259**
BSI					0.768***	0.957***	-0.104	0.040
$MSI^{\perp}$						0.802***	0.000	0.199*
$BSI^{\perp}$							0.000	0.051
$RMRF_t$								0.127

Table 4
Correlations between sentiment indices measured on a monthly basis over the period 1999-2006, namely the French sentiment index (FSI), the two Baker-Wurgler indices (BW1 and BW2), our market sentiment index (MSI) and the Buy-Sell imbalance index (BSI). The two last columns give the correlations between sentiment indices and market return. Subscript t denotes contemporaneous correlations calculated over 96 months, and subscript $t+1$ means lagged correlations. In this case they are calculated with a 95-month vector of sentiment indices between January 1999 and November 2006 and the market return is a 95-month vector covering February 1999 to December 2006. Of course, the contemporaneous correlation is zero for the two orthogonalized indices.

	FSI	$BW1$	$BW2$	MSI	BSI	$MSI^{\perp}$	$BSI^{\perp}$
Panel A: Equation 13 without controls							
b	-0.001	-0.007	-0.006	-0.054^{**}	-0.145^{*}	-0.068^{***}	-0.154^{**}
t-stat	-1.324	-1.365	-1.417	-2.069	-1.95	-2.784	-2.06
p-value	0.189	0.176	0.160	0.041	0.054	0.007	0.041
$\overline{R}^2$	0.008	0.006	0.005	0.071	0.054	0.102	0.056
Panel B: Equation 14 with controls							
d	-0.001^{**}	-0.011^{***}	-0.011^{***}	-0.035	-0.14^{**}	-0.054^{**}	-0.145^{**}
t-stat	-2.09	-2.97	-2.74	-1.44	-2.29	-2.378	-2.24
p-value	0.040	0.004	0.007	0.153	0.024	0.019	0.027
$\overline{R}^2$	0.220	0.225	0.224	0.211	0.238	0.246	0.236

Table 5

Coefficients of sentiment when regressing the returns of a long-short portfolio based on size, on sentiment measures (with Newey-West consistent estimates). Panel A gives the coefficient of sentiment in the simple regression:

$$R_{Smallcaps,t} - R_{Bigcaps,t} = a + bSENTIMENT_{t-1} + \varepsilon_t$$

Panel B provides the same coefficient when controlling for Fama-French factors and Carhart momentum factor. The regression equation is:

$$R_{Smallcaps,t} - R_{Bigcaps,t} = c + dSENTIMENT_{t-1} + \beta RMRF_t + hHML_t + mMOM_t + \varepsilon_t$$

The sentiment measures are the French sentiment index (FSI), the two Baker-Wurgler indices ($BW1$ and $BW2$), the market sentiment (buy-sell imbalance) index $MSI(BSI)$ and the corresponding orthogonalized versions $MSI^{\perp}$ and $BSI^{\perp}$. When sentiment is not considered in the controlled equation, the adjusted R^2 of the regression is 0.188.

	Eq. 13 without control		Eq. 14 with control	
K = 10	MSI	$MSI^{\perp}$	MSI	$MSI^{\perp}$
b	-0.052^{*}	-0.073^{***}	-0.034	-0.06^{***}
t-stat	-1.93	-3.02	-1.43	-2.77
p-value	0.057	0.003	0.157	0.007
$\overline{R}^2$	0.061	0.105	0.208	0.253
K = 20	MSI	$MSI^{\perp}$	MSI	$MSI^{\perp}$
d	-0.054^{**}	-0.068^{***}	-0.035	-0.054^{**}
t-stat	-2.069	-2.784	-1.442	-2.378
p-value	0.041	0.007	0.153	0.019
$\overline{R}^2$	0.071	0.102	0.211	0.246
K = 30	MSI	$MSI^{\perp}$	MSI	$MSI^{\perp}$
d	-0.054^{**}	-0.064^{**}	-0.035	-0.049^{**}
t-stat	-1.99	-2.99	-1.33	-2.018
p-value	0.05	0.014	0.186	0.047
$\overline{R}^2$	0.068	0.089	0.209	0.236

Table 6
Coefficients of sentiment when regressing the returns of a long-short portfolio based on size on sentiment measures MSI and MSI orthogonalized, with Newey-West consistent estimates. 3 values of K varying from 10 to 30 by steps of 10 are considered. The two left columns give the regression coefficients of the equation without control variables. The two last columns provide the coefficients when the Fama-French and Carhart factors are introduced as control variables

	Without control		With control	
$W > 100$	MSI	$MSI^{\perp}$	MSI	$MSI^{\perp}$
b	-0.054^{**}	-0.067^{***}	-0.035	-0.054^{**}
t-stat	-2.06	-2.78	-1.43	-2.38
p-value	0.041	0.006	0.153	0.019
$\overline{R}^2$	0.081	0.102	0.211	0.246
$W > 1,000$	MSI	$MSI^{\perp}$	MSI	$MSI^{\perp}$
d	-0.056^{**}	-0.070^{***}	-0.038	-0.056^{**}
t-stat	-2.10	-2.88	-1.53	-2.49
p-value	0.038	0.005	0.13	0.015
$\overline{R}^2$	0.072	0.105	0.214	0.249
$W > 5,000$	MSI	$MSI^{\perp}$	MSI	$MSI^{\perp}$
d	-0.059^{**}	-0.073^{***}	-0.043^{*}	-0.06^{***}
t-stat	-2.16	-2.96	-1.74	-2.63
p-value	0.033	0.004	0.085	0.01
$\overline{R}^2$	0.071	0.103	0.219	0.252

Table 7
Coefficients of sentiment when regressing the returns of a long-short portfolio based on size on sentiment measures MSI and MSI orthogonalized, with Newey-West consistent estimates. 3 minimum portfolio values, respectively 100, 1,000 and 5,000 euros are considered. The two left columns give the regression coefficients of the equation without control variables. The two last columns provide the coefficients when the Fama-French and Carhart factors are introduced as control variables. In this case, introducing no sentiment measure leads to a determination coefficient of 0.188

Appendix: Sentiment indices

We briefly present the French sentiment index and the two Baker-Wurgler indices and give references for more detailed presentations.

A. The French consumer sentiment index

It is based on the same principles as the ICS. The French Institute of statistics realizes a monthly phone survey[12] with around 2,000 households. It also provides information along several dimensions linked to perception of economic conditions and expectations. The results are presented as differences between good and bad opinions for each dimension. The synthetic index used in this paper is based on the following indicators[13]:

1) Past personal financial situation

2) Expectation about future evolution of financial situation

3) Opportunity to invest in consumption goods

4) Past standard of living

5) Expectation about future evolution of standard of living

6) Unemployment perspectives

7) Saving capacity

The synthetic measure is obtained through a factor analysis. There also exists a summary index which is like the ICS, an arithmetic average of the items 1 to 5.

B. The Baker-Wurgler indices

Baker and Wurgler (2006) argue that it is difficult to rely on a unique variable to represent sentiment. They build two sentiment indices as linear combinations of the six following variables:

1) The closed-end fund discount

[12] The methodological details are explained at http://www.insee.fr/fr/indicateurs/ind20/method_idconj_20.pdf
[13] The complete questionnaire can be found at http://www.bdm.insee.fr/bdm2/
documentationGroupe.action?
codeGroupe=389

2) The natural logarithm of the NYSE share turnover ratio (detrended by the 5-year moving average)

3) The number of IPOs

4) The average first-day return on IPOs

5) The share of equity issues in total equity and debt issues

6) The dividend premium (log difference of the average market-to-book ratios of dividend payers and non payers)

The authors define their first index as the loadings on the first principal component in a PCA of the 6 variables. One possible criticism of this index (mentioned by Baker and Wurgler) is that it is difficult to disentangle what comes from sentiment and what is driven by the business cycle in the loadings. Consequently, they build a second index with the same approach, except that the six variables are now the residuals of the regression of the initial variables on growth in the industrial index, growth in consumer durables, non durables and services and a dummy variable for NBER recessions (see Baker and Wurgler, 2006, p 1657).

Working Papers

Laboratoire de Recherche en Gestion & Economie

http://ifs.u-strasbg.fr/large/publications/publications.htm

Université de Strasbourg
Pôle Européen de Gestion et d'Economie
61 avenue de la Forêt Noire
67085 Strasbourg Cedex

http://ifs.unistra.fr/large