Laboratoire
de Recherche
en Gestion
& Economie

Working Paper
2012-04

In search of positive skewness: the case of individual investors

Patrick Roger, Marie-Hélène Broihanne, Maxime Merli

March 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

Institut de Finance de Strasbourg

In search of positive skewness: the case of individual investors

Patrick ROGER,[*] Marie-Hélène BROIHANNE, Maxime MERLI[†]

March 2012

Abstract

In this paper, we first prove analytically that the skewness of returns of portfolios built with Arrow-Debreu securities decreases with diversification. Through simulations, we also show that this result remains true in a financial market with a finite number of states of nature. We then analyze the behavior of over 85,000 individual investors at a large brokerage house. Though the main determinant of underdiversification is the portfolio value we find that the skewness of returns remains significant in explaining diversification after controlling for this value. Moreover, we show that the decrease in skewness induced by diversification is essentially driven by the share of total variance of stock returns due to common factors. These findings extend those of Mitton and Vorkink (2007) and explain the variability over time of the relationship between skewness and diversification.

Keywords: Underdiversification, skewness, individual investors, **JEL**: G11, G12

[*]Corresponding author, LARGE Research Center, EM Strasbourg Business School, University of Strasbourg, 61 avenue de la forêt noire, 67085 Strasbourg CEDEX, France. E-mail: proger@unistra.fr, Phone number: 0033368852156.

[†]LARGE Research Center, EM Strasbourg Business School, University of Strasbourg, 61 avenue de la forêt noire, 67085 Strasbourg CEDEX, France. We thank Laurent Deville, Gunter Franke, Jens Jackwerth, Gregory Nini, Charles Noussair, Winfried Pohlmeier, Pierre Six, Marc Willinger, the participants of the DMM meeting (2011, Montpellier), the 28th international AFFI Conference (2011, Montpellier), the Konztanz-Strasbourg Workshop (2011, Königsfeld) and the Behavioral Insurance Workshop (2011, München) for comments and suggestions. The financial support of the European Savings Institute (OEE, Observatoire de l'Epargne Européenne) is gratefully acknowledged. We thank Tristan Roger for valuable research assistance and computer programming. E-mails: mhb@unistra.fr, merli@unistra.fr.

In an economy governed by the standard theory of portfolio choice (Markowitz, 1952), investors hold diversified portfolios. These investors seek returns with a high first moment and a low second moment. The actual economy presents a substantially different reality. Lease et al. (1974) and Blume and Friend (1975), followed by Kelly (1995), were the first to highlight the underdiversification of portfolios held by retail investors. More recently, a number of empirical studies (Odean 1999, Kumar 2007, Mitton and Vorkink 2007, Goetzman and Kumar 2008), using large datasets, have shown that individual investors hold largely underdiversified portfolios, which contain fewer than five stocks on average[1].

In the theoretical literature, several complementary explanations are provided for underdiversification. Some explanations are linked to skewness seeking by investors (Brunnermeier and Parker, 2005, Brunnermeier et al., 2007, Mitton and Vorkink 2007, Barberis and Huang 2008), whereas others are based on the existence of solvency constraints for investors and/or on the existence of market imperfections, such as transaction costs (Liu 2010).

The first contribution of this paper is theoretical. In an Arrow-Debreu economy with a finite number of states of nature, we prove that the skewness of returns of equally weighted portfolios of pure contingent claims decreases when the number of different securities in the portfolio increases[2]. Considering the first measure of diversification used by Mitton and Vorkink (2007), that is, the inverse of the number of stocks held, we show that the variance of portfolio returns is a linear function of this measure and that the third central moment is a quadratic function of the same measure. Through simulations, we show that this result generally remains true (in a finite-state space) for stock portfolios[3].

Our second contribution is empirical: through a detailed analysis of the trading records and portfolios of 87,373 individual investors, we show that return skewness is an important explanatory

[1] Calvet *et al.* (2007) obtained similar results for Sweden with the exception that Swedish investors appear to have slightly more diversified portfolios than U.S. investors.

[2] Simkowitz and Beedles (1978) and Conine and Tamarkin (1981) already note that diversification may decrease the skewness of returns.

[3] Harvey and Siddique (1999, 2000) and Chen *et al.* (2001) show that the average skewness of single stocks is positive in most periods and that the market skewness is negative most of the time. More recently, Albuquerque (2011) obtained the same results except for the second half of 1987 (due to Black Monday). The skewness of the equally weighted market portfolio is negative 77% of the time.

variable of diversification choices, even after controlling for portfolio value. Mitton and Vorkink (2007) obtained the same result but did so without controlling for portfolio value. In fact, portfolio value has been shown to be the main determinant of diversification choices, which supports Liu's model which is based on the assumption that retail investors are subject to solvency constraints. Liu shows that the number of stocks in one's portfolio is an increasing function of the amount to be invested in the stock market. However, in his model, only the first moment of the probability distribution of returns is important for single-stock owners. As mentioned above, our results indicate that the skewness of returns is also important for diversification choices.

Our third contribution aims to explain why the link between the decrease in skewness due to diversification is highly variable over time. For example, Mitton and Vorkink (2007) find a difference in skewness between single stocks and diversified portfolios equal to 0.2 in January 1991, 0.36 in January 1993 and 0.54 in January 1996. We show that this variability is essentially due to the evolution of the aggregate share of variance due to common factors or, more simply, to the average correlation of stock returns. In a multi-factor model à la Chamberlain and Rotschild (1983) with, for example, K common factors, the reduction in skewness due to diversification is a decreasing function of the sum of the K first eigenvalues of the covariance matrix of returns (in percentage of the trace). We first show this relationship in simulated markets and confirm it on our data. This important result shows that in bear markets (where market volatility _ or volatility due to common factors _ is higher), underdiversification does not significantly increase skewness. However, diversification does not substantially reduce the portfolio variance because systematic risk is the most important component of the total risk during these periods.

The paper is organized as follows. The first section reviews the literature on portfolios' (under)diversification and skewness seeking by investors. In section 2, we present the analytical and simulation results in Arrow-Debreu markets. Section 3 describes our database and section 4 details our empirical results. The last section concludes the paper.

1 Related literature

The search of positive skewness in portfolio returns may be justified by two psychological traits. First, the propensity to gamble may lead to the preference for positively skewed portfolios in the hope of gaining a high return, even with a very low probability. Investors with such preferences prefer lottery-type stocks with low prices, high idiosyncratic volatility and high skewness (Kumar 2009, Bali, Cakici and Whitelaw 2011). Meanwhile, preferring high skewness may simply reveal prudent behavior in the manner described by Kimball (1990). In this case, investors are said to be downside risk averse (Menezes et al. 1980, Eeckhoudt and Schlesinger, 2006), and their utility function is mainly characterized by a positive third-order derivative. Several recent experimental studies show that approximately 60% of participants are prudent when faced with lottery choices. Their decision process is therefore compatible with positive skewness seeking (Tarazona-Gomes 2004, Deck and Schlesinger 2010, Ebert and Wiesen 2011). If underdiversification is a means of capturing high skewness, then investors who choose their optimal portfolios, considering not only the first two moments of the distribution of returns but also the third one, will build underdiversified portfolios. In fact, one of the first attempts to introduce the third moment of the distribution of returns in a portfolio choice model was proposed by Kraus and Litzenberger (1976), followed by Harvey and Siddique (2000), who provided empirical support for this model. More recently, Mitton and Vorkink (2007) also based their theoretical analysis on a three-moment model and show that heterogeneity in the preference for skewness induces underdiversification in equilibrium.

In the framework of non-expected utility models, the desirability of positive skewness appears to occur because investors distort probabilities. Shefrin and Statman (2000), in their behavioral portfolio theory, consider that investors' decisions are driven by a mix of hope and fear. Hope (fear) tends to transform probabilities in an optimistic (pessimistic) way. According to this approach, optimal portfolios are, roughly speaking, composed of a risk-free asset (motivated by fear) combined with a lottery ticket (the hope to become rich). Such a portfolio is obviously positively skewed.

Barberis and Huang (2008) assume that investors obey Cumulative Prospect Theory (Tversky and Kahneman 1992). Their utility functions are concave for gains and convex for losses. Moreover, they distort probabilities in such a way that extreme outcomes are overweighted. When a positively skewed asset is traded on the market, it becomes overpriced because of the overweighting of the largest positive outcomes. Cumulative Prospect Theory (CPT) overweights the two tails of the distribution of outcomes; it leads investors to avoid large losses and to seek large gains, even if the corresponding objective probability is very low. Again, the result is that positive skewness is attractive for such investors.

One potential drawback of CPT is that the transformation of probabilities is "exogenous" in the sense that it depends solely on the ranking of outcomes, not on their values. Consider, for example, two successive draws of a state lottery, and assume that the jackpot is not hit at the first draw. At the second draw, the potential outcomes are almost identical, except for the jackpot, which has increased between the two draws because of the rollover. An agent obeying CPT does not change the weights of the outcomes because the objective probability of hitting the jackpot is the same and this event is still ranked first.

Consequently, Brunnermeier and Parker (2005) go one step further by introducing the distorted probability measure as a decision variable, referring to optimal expectations or optimal beliefs. They consider a forward-looking investor who, when making a decision, maximizes the average of her current (anticipatory) utility and her expected future utility. The anticipatory utility is higher when the investor is optimistic about future prospects. This distortion of beliefs leads to suboptimal investment decisions in terms of resource allocation and portfolio choice. Nevertheless, these authors show that a slightly optimistic change in beliefs generates a first-order gain in current utility but only a second-order loss in future utility due to suboptimal investment decisions. In the same vein, Gollier (2005) shows that optimal expectations correspond to beliefs focusing on the best and the worst state.

It is then not optimal for such agents to select a portfolio under the real probability measure (as rational agents do). Brunnermeier et al. (2007), building on Brunnermeier and Parker (2005), elaborate a simple model of a complete market of Arrow-Debreu securities and conclude that the probability of one state is overvalued while the probabilities of the other states are undervalued by such investors. They obtain an optimal portfolio that has the same shape as that found by Shefrin and Statman (2000), that is, a risk-free asset combined with a positively skewed asset (equivalent to a lottery ticket).

The attractiveness of positive skewness in returns can also be caused by a type of "jackpot effect", such as in-state lotteries. It is now well documented that the demand for state lotteries is essentially determined by the jackpot size, meaning that players are attracted by the best outcome, even if the corresponding objective probability of occurrence is infinitesimal (Cook and Clotfelter 1993, Garrett and Sobel 1999, Walker and Young, 2001, Forrest et al. 2002). In stock markets, this effect has been recently illustrated by Kumar (2009) and Bali, Cakici and Whitelaw (2010). Kumar shows that for some categories of individual investors, there is a strong link between portfolio choice and behavior in gambling markets such as state lotteries. More precisely, those who are prone to betting on state lotteries are also prone to choosing low-priced stocks with high idiosyncratic risk and high positive skewness. Bali, Cakici and Whitelaw (2010) do not analyze the behavior of individual investors, instead ranking stocks according to their maximum one-day return over the previous month. They find that future returns are a decreasing function of this one-day maximum return. In other words, lottery-like stocks are overpriced. They also show the persistence of this ranking over time by calculating transition probability matrices from one month to the next. A total of 35% of stocks in the highest decile for one month are in the same decile the following month.

Whatever the interpretation, the credo of diversification becomes at stake when preference for positive skewness is introduced in the decision process. In fact, diversification reduces both (undesirable) idiosyncratic volatility and (desirable) positive skewness.

2 Skewness and diversification: the case of Arrow-Debreu markets

In this section, we start by developing analytical formulas for the first three moments of equally weighted portfolios of Arrow-Debreu (henceforth, AD) securities and link these results to usual measures of diversification. More precisely, when diversification is measured by the inverse of the number of stocks in portfolios, the variance of portfolio returns is a linear function of the diversification index, and the third central moment of returns is a quadratic function of the diversification measure. In the next subsection, simulations provide a means of extending this result to more general frameworks.

2.1 Moments of portfolios of Arrow-debreu securities (AD securities)

Let $\Omega = \{\omega_1, ..., \omega_n\}$ denote a finite state-space with n equally-likely states of nature and assume that all Arrow-Debreu securities, denoted $X_1, ...X_n$, are traded. X_i pays 1 in state ω_i and 0 elsewhere. $(p_1, ..., p_n)$ stands for a sequence of equally-weighted portfolios containing respectively $1, 2...n$, AD securities. Without loss of generality, we assume that p_k contains $1/k$ units of each of the first k securities[4].

Before analyzing portfolios, we briefly recall the elementary properties of the moments of AD securities.

Proposition 1 *For any* $1 \leq k \leq n$,

$$\begin{aligned} E(X_k) &= m_k = \frac{1}{n} &(1)\\ V(X_k) &= \frac{n-1}{n^2} &(2)\\ E\left[(X_k - m_k)^3\right] &= \frac{(n-1)(n-2)}{n^3} &(3)\\ cov(X_k, X_{k^*}) &= -\frac{1}{n^2} \text{ if } k \neq k^* &(4) \end{aligned}$$

Proof. See the Appendix ■

Consider now a portfolio p_k invested in the first k AD securities and denote μ_k (σ_k^2) the expectation (variance) of payoffs of p_k.

[4] Because states are equally likely, there is no reason to consider different prices for AD securities. Investing $1/k$ in each of the first k securities then generates a cost independent of k.

Proposition 2

$$\begin{aligned}\forall k, \mu_k &= \frac{1}{n} \\ \sigma_k^2 &= \frac{1}{n}\left(\frac{1}{k} - \frac{1}{n}\right)\end{aligned} \tag{5}$$

Proof. See the Appendix ■

The inverse of the number of stocks in portfolios is often considered a measure of diversification (denoted D_1 by Mitton and Vorkink (2007)). Proposition 2 shows that the variance of returns increases linearly with D_1. When $k = n$, the portfolio is risk-free and the variance of payoffs is equal to 0.

Denote now s_k^3 the third central moment of p_k defined by:

$$s_k^3 = E((\frac{1}{k}\sum_{j=1}^{k} X_j - \mu_k)^3) = \frac{1}{k^3}E((\sum_{j=1}^{k} X_j - \frac{k}{n})^3) \tag{6}$$

Denoting $Y_k = \sum_{j=1}^{k} X_j$ gives $s_k^3 = \frac{1}{k^3}E((Y_k - \frac{k}{n})^3)$. The specificities of AD securities imply that $E(Y_k^3) = E(Y_k^2) = \frac{k}{n}$. In fact, these relations simply come from the fact that $X_j^m X_{j^*}^t = 0$ for any pair (m, t) of strictly positive integers and different indices j and j^*. We now get easily s_k^3.

Proposition 3 *The central third moment of p_k is valued:*

$$s_k^3 = \frac{1}{n^3}\left[\left(\frac{n}{k} - 1\right)\left(\frac{n}{k} - 2\right)\right] \tag{7}$$

Proof. See the Appendix ■

We know that $k < n$; consequently an equally weighted portfolio has a positive skewness as long as the number of AD securities that it contains is lower than $n/2$. Beyond this threshold, skewness becomes negative. When n is even, the distribution of returns is symmetric for $k = n/2$, leading to a zero skewness for the portfolio return. Using the above diversification measure D_1, we find that the third-order moment increases quadratically in D_1. In fact, we have the following:

$$s_k^3 = \frac{1}{n}\left[\left(D_1 - \frac{1}{n}\right)\left(D_1 - \frac{2}{n}\right)\right] \tag{8}$$

We can also establish a very simple relationship between s_k^3 and σ_k^2 using equations 5 and 7.

$$s_k^3 = \sigma_k^2\left(D_1 - \frac{2}{n}\right) \tag{9}$$

Insert Table 1 around here

Table 1 shows the evolution of variance and third-central moment as a function of the number of AD securities in portfolios when $n = 20$. The third moment becomes slightly negative when the number of AD securities is greater than 10 but the variance decreases at a faster rate. This finding implies that standardized skewness, defined as s_k^3/σ_k^3 becomes largely negative when the portfolio is sufficiently diversified (see Figure 1). This remark is in line with the abovementioned positive skewness of single-stock returns and negative skewness of highly diversified portfolios. Positive skewness observed for single stocks (the most common case) is often explained by an overreaction to good news and underreaction to bad news (Nagel, 2005, Xu, 2007). When estimating the skewness of a single-stock with a time-series of returns, it is common to find sequences of returns with one or a few very high values due to overreaction and a number of low or moderate values due to underreaction. This shape leads to a positive estimation of skewness. When considering the time-series of returns of a diversified portfolio, isolated high values are less likely because good news does not come at once for all stocks in the portfolio. On the contrary, it is well documented that correlations of stock returns increase in hard times[5]. Consequently, low returns are more likely to be observed simultaneously, leading to negative skewness for portfolio returns. A "similar" phenomenon is observed with portfolio of Arrow-Debreu securities. The payoff of a single AD is typically positively skewed, paying 1 in one state and 0 otherwise. An equally weighted portfolio

[5] For international equity returns, Longin and Solnik (2001) show that the asymmetry of correlations is statistically significant. Campbell et al. (2002) and Ang and Bekaert (2002, 2004) also identify asymmetric correlation between bull and bear regimes, higher correlations appearing in the bear regime and lower correlations in the bull regime.

comprising a large number (say k) of AD securities pays $1/k$ on k states and 0 on the remaining $n-k$ states. When k approaches n, the portfolio payoff becomes negatively skewed because losses are less likely but larger (for a given initial investment).

Insert Figure 1 around here

2.2 Simulations

2.2.1 Portfolios of Arrow-Debreu securities

The preceding section assumes equal weights in portfolios, allowing simple analytical results to be obtained concerning the evolution of skewness as a function of the number of Arrow-Debreu securities. To determine whether this finding remains true in a more general framework, we simulate portfolio weights in the following way. Still considering 20 states of nature, we randomly select, for each number m between 2 and 19, 1000 portfolios of m securities. For a given m-security portfolio, we draw m random numbers $x_1, ..., x_m$ between 0 and 100 and define the weights as $w_k = x_k / \sum_{j=1}^{m} x_j$. We consider only positive weights since individual investors almost never engage in short-selling.

Insert Figure 2 around here

For each m, we calculate the average skewness and the 99% confidence bounds (5th lowest and 5th highest skewness in a sample of 1000 portfolios).

Figure 2 shows the evolution of the average skewness and the confidence bounds as well (dashed lines). Of course, there is no uncertainty for single-stock portfolios because of the analytical solution given in the proposition 1. The average skewness of simulated portfolios is always positive regardless of the number of stocks in the portfolios. It approaches 0 only for the maximum number of stocks. At first glance, this finding could seem surprising because in the equally weighted case of the preceding section, the skewness was negative when $k > n/2$. The reason for this phenomenon is that portfolio weights are simulated according to a uniform multidimensional distribution. It is revealed that the proportion of portfolios close to the equally weighted case is very low and that these are the portfolios with the lowest skewness. The bold curve in figure 2 is comparable to the results obtained by Mitton and Vorkink (2007, table 3, p. 1271), that is, it depicts a similar evolution of skewness with respect to diversification. Consequently, it appears that underdiversification is an effective means of capturing high skewness. It is not even necessary to pick highly skewed stocks to achieve a highly skewed portfolio. There is a "mechanical" relationship between diversification and return skewness.

2.2.2 The general case

In a finite state-space, a stock is a portfolio of Arrow-Debreu securities. To generalize the above results, we now consider n stocks traded on the n-state market. To generate the payoffs of stock k, we first draw at random a number n_k of AD-securities in the set of n assets. We then define the payoffs of stock k by drawing n_k random weights (summing to one), as in the preceding section. We therefore build a (n, n) matrix of payoffs of single stocks, which corresponds to one simulated market.

Taking a geometric approach to the problem is useful to understand the link between diversification and skewness. In the AD market, adding one AD security entails adding one dimension to the space spanned by securities because AD securities are orthogonal vectors. However, usual stocks, that is, portfolios of AD securities, may be correlated, and the dimension spanned by stocks may be lower than the number of states. The probabilistic view of the same story is to say that when market (idiosyncratic) risk represents a large (small) share of total variance, stocks do not span a large space and diversification will not substantially decrease the skewness of portfolio returns (compared to the skewness of single stocks).

Simulation is an interesting and simple way to determine whether our conjecture is true. Remember that empirical studies show that skewness decreases with diversification but that the steepness of the slope changes over time. Mitton and Vorkink (2007) find a skewness difference between single-stock portfolios and highly diversified portfolios equal to .2 in January 1991, .36 in January 1993 and, finally, .54 in January 1996.

To introduce the methodology, consider the simple one-factor CAPM-like model. The variance of returns of a given stock k is equal to $\beta_k^2\sigma_M^2 + \sigma(\varepsilon_k)^2$ where σ_M^2 is the variance of the market portfolio, β_k is the β of stock k and $\sigma(\varepsilon_k)^2$ is the idiosyncratic variance of stock k. An estimate of $\sigma_M^2 \sum_{k=1}^n \beta_k^2 / \sum_{k=1}^n \sigma_k^2$ is given by the first eigenvalue of the covariance matrix of returns (as a percentage of the trace of the covariance matrix which is $\sum_{k=1}^n \sigma_k^2$).

More generally, in an m-factor model, the variance due to common factors is the sum of the m first eigenvalues (as a percentage of $\sum_{k=1}^n \sigma_k^2$). Our conjecture is then that the average skewness decreases faster with diversification when the variance due to common factors is low. Consequently, the rank correlation between the decrease in skewness and the share of variance due to common factors should be significantly negative.

It is well-known that estimating the number of common factors by using a principal component analysis (as suggested by Chamberlain and Rotschild, 1983) is difficult because there is a bias towards the identification of a single-factor model in finite samples (Harding, 2007). To take this bias into account, we allow up to $m = 5$ common factors in the analysis.

For a given simulated market, we measure the decrease in skewness due to diversification by the difference between the average skewness of single stocks and the skewness of an equally weighted portfolio of all stocks.

Insert Table 2 around here

Table 2 summarizes the results. Panel A (B, C) gives the main results when there are 20 (60, 100) assets traded on the market and the same number of states of nature. In each panel, the first line displays the average cumulated percentage of variance for the m first factors where $m = 1, .., 5$. The second line provides the Spearman rank correlation between this percentage of variance and the decrease in skewness. Naturally, because these correlations are valued over 1,000 simulations, they are significant at all the standard levels. For example, in panel A, the first factor represents (on average) 24.06% of the global variance. If stock returns were independent, this figure would be 5% with 20 states of nature. The rank correlation in the one-factor model is equal to -0.469 and is the correlation between the vector of decreases in skewness and the vector of percentages of "market" variance. The other figures in the table are interpreted in the same way.

These results show that if underdiversification generates skewness, a given level of diversification is likely to provide different levels of skewness depending on the market conditions. In particular, we mentioned earlier that bearish markets tend to increase the average correlation between stock returns (Longin and Solnik, 2001, Campbell et al., 2002, Ang and Bekaert, 2002, 2004), meaning that the percentage of variance linked to common factors is higher. We then observe a lower decrease in skewness due to diversification. In the next section, we illustrate this point using our large sample of individual investors. In fact, a low percentage of simulations leads to an increase in skewness (6.6%, 2.4% and 2.7% in panels A, B and C, respectively) when the number of stocks in the portfolio increases, meaning that the equally weighted portfolio is more positively skewed than the average stock. These cases correspond to high levels of market variance. For these situations, 30.1% (20.03%, 16.8%) of variance stems from the first factor, compared to an average of 24.06% (14.27%, 11.79%) in table 2.

3 Data and descriptive statistics

3.1 Investors data

The 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, nine million trades made by 92,603 investors. The file of trades combines the following information for each trade: the 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 into and exit from the database, opening and/or closing dates of all accounts and region of residence.

Some investors open an account within this period, while some others close their account before the end of the period. As it would make no sense to analyze portfolios by the 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 continued trading later. 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. A total of 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 3 shows three time-series. The upper dotted curve represents the average number of stocks held by investors ($\times 10^4$). The value of this curve varies from 5.5 to 6.8, and the median is 3 or 4 for the entire period. The difference between median and mean is explained by a low percentage of investors holding largely diversified portfolios with several hundred stocks. These figures show that it is reasonable to postulate that individual investors hold underdiversified portfolios. The striking feature of this curve is the sharp increase in the average number of stocks just before the dotcom bubble burst, that is, during the first months of 2000. A decrease to the former level of diversification is observed following this event. For the remainder of the period, the average number of stocks is roughly stable. The evolution of the number of stocks is different from that observed by Goetzman and Kumar (2008) on a sample of U.S. investors. As mentioned before, the authors found an increase in diversification over the period 1991-1996 because the market was bullish for almost the entire period.

The middle bold curve and the bottom dashed curves provide the evolution of the mean and median portfolio values, respectively. Starting in the first month of 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 first 15 months, up to the bursting of the Internet bubble 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 average portfolio values in our sample does not differ from the evolution of the stock market.

[**Insert Figure 3 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. In fact, a few investors are very wealthy, compared to the average investor; they move upward the average portfolio value in a significant way. It is worth to mention that 0.2% of investors hold a stock portfolio worth over one million euros.

Table 3 provides some additional detailed statistics at three points in time, January 2000, January 2003 and January 2006[6]. We adopt the same presentation used in Table 2 of Mitton-Vorkink (2007). At the end of each month, we divide investors into seven categories (first column of Table 3). 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 provides the number of investors in each category. The last four columns provide summary statistics about portfolio values, namely, the mean, the first quartile, the median and the third quartile. There is a large proportion (approximately 20%) of single-stock owners, and in all categories, the mean portfolio value is much higher than the median, even among single-stock owners. This finding reinforces the preceding remark about figure 3. In most cases, the mean is close to the third quartile.

[**Insert Table 3 around here**]

The market activity of investors in our sample is also highly variable over time. Figure 4 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 found in the two types of trades up to April 2000. Approximately 110,000 monthly buy trades were realized in February, March and April 2000. An equivalent decrease is then observed up to September 2001. Of course, even though the French market remained open after 9/11, the volume was considerably lower that month.

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

[**Insert Figure 4 around here**]

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

3.2 Stock data

Price data come from two sources, Eurofidai for stocks traded on Euronext[7] and Bloomberg for the other stocks. We used daily prices to estimate the moments of the distribution of returns on stocks and investors' portfolios. In our sample, the investment universe contains 2,491 stocks, meaning that each of these stocks has been traded at least once over the sample period. There are 1,191 French stocks, the remaining coming from all over the world but essentially from the U.S. (1,020 stocks), the United Kingdom (62), the 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 over 90% of the trading volume, as shown in panel A of table 4 which illustrates the well-known home bias puzzle[8]. This phenomenon represents the reason why most comparisons in this paper are related to the French market. Moreover, we observe that despite the large number of U.S. stocks, the trading volume for these stocks is very low. 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 4 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. A total of 96.97% of these investors held French stocks (meaning that approximately 3% held only foreign stocks), 21.05% held Dutch stocks and only 3.97% held U.S. stocks, despite the large number of U.S. stocks in the database (that is, stocks traded at least once over the period).

[Insert Table 4 here]

3.3 Measures of diversification and skewness

Despite our theoretical results on Arrow-Debreu markets in section 2, it is unclear (in real markets) whether underdiversified portfolios should bear more skewness. To answer this question we first

[7] http://www.eurofidai.org. Part of this database has been recently used by Foucault *et al.* (2011) in their study of retail trading and volatility in the French market and by Baker *et al.* (2012) to study the contagion of sentiment across countries, including France and the U.S.

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

apply the methodology of Mitton and Vorkink (2007) to our sample. Using two portfolio diversification measures, we rank the individual portfolios according to these measures and calculate the mean and median skewness within each decile. If underdiversification is caused by skewness seeking, skewness should be high for deciles of less diversified portfolios. We perform these calculations for each quarter, skewness being estimated with one quarter of daily returns[9]. We first use returns subsequent to the portfolio formation date, as the realized skewness for a given portfolio is the skewness of future returns. However, from a behavioral point of view, investors can base their choices on the observation of past skewness. We then assess whether such a relationship also appears between past skewness and diversification.

The two diversification measures, denoted D_1 and D_2, are defined as follows. D_1 is simply the inverse of the number of different stocks in the portfolio.

$$D_1^j = \frac{1}{n_j} \tag{10}$$

where n_j is the number of stocks in investor j's portfolio. A low value of D_1 is then associated with a high level of diversification.

Though simple, this measure does not take into account the weighting of securities within portfolios. Consequently, we introduce as D_2 the Herfindahl index of the weights of securities in the investor's portfolio.

$$D_2^j = \sum_{i=1}^{n} w_{ij}^2 \tag{11}$$

where w_{ij} is the weight of security i in investor j's portfolio. D_2 also takes higher values for lower levels of diversification. The two measures are then positively correlated.

To measure the standardized skewness of portfolio returns, we use the usual estimate with one quarter of daily returns

$$\widehat{S_k}^3 = \frac{\frac{1}{n}\sum_{t=1}^{n}(r_t - \overline{r})^3}{\widehat{\sigma}^3} \tag{12}$$

where $\overline{r}$ is the average daily return and $\widehat{\sigma}^3$ the cube of the estimated standard deviation of daily

[9]The results (not reported here) are almost identical when considering one year of daily returns.

returns. One advantage of equation (12) is that it is standardized by variance (or standard deviation). This equation represents a way to take into account the mechanical positive link between variance and skewness that is illustrated in section 2. However, in an expected utility framework, the utility of random wealth is linked to the third moment of returns, not to the standardized skewness.

4 Empirical results

In this section, we first perform an investor-level analysis. We analyze the link between underdiversification and skewness in returns and confirm that the mechanical link obtained through simulations also appears in real data. We develop our study by showing that the decrease in skewness due to diversification is strongly driven by market conditions and, more precisely, by the sharing of global variance between systematic and idiosyncratic variance. In other words, portfolio skewness is essentially linked to the number of stocks in the portfolio and not as much to which stocks are chosen to build the portfolio. To reinforce this conclusion, we use the approach of Mitton and Vorkink (2007) to perform a stock-level analysis aiming to answer the following question: do underdiversified investors concentrate on highly skewed stocks? In contrast to these authors, we do not obtain a clear answer. The strength of the relationship is shown to depend on the sharing of global variance between systematic and idiosyncratic variance.

Finally, through a panel data analysis, we show that the main determinant of diversification choices is the amount to be invested in the portfolio. Nevertheless, return skewness remains a significant explanatory variable of diversification choices after controlling for portfolio value.

4.1 Diversification and portfolio skewness

To study the link between diversification and skewness, we sort all investor portfolios into deciles according to measures of diversification D_1 and D_2 at the beginning of each quarter. We calculate the average value of $\widehat{S_k}^3$ within each decile, as well as the average returns and standard deviations (which are annualized in table 5). We cannot present the results for all quarters here so we take the same three points in time as before[10].

[Insert Table 5 here]

Table 5 provides the estimates made when subsequent returns are used to calculate moments. Concerning D_1, the number of investors differs between deciles because D_1 is a discrete variable (the inverse of the number of stocks). It is revealed that it is impossible to arbitrarily allocate investors in different deciles when they have the same value for D_1. Concerning the Herfindahl index D_2, this problem concerns only the single-stock holders. A consequence is that the number of investors is the same for the two indices in the last decile (single-stock holders).

Whatever the diversification index under consideration, there is a stable relationship between D_i, $i = 1, 2$ and the second and third moments of the distribution of returns. A higher D indicates lower diversification, higher variance and higher skewness. This relationship is consistent with our theoretical results of section 2 and is in line with the empirical results of Mitton and Vorkink (2007). In table 5 we adopt a presentation similar to their table 3 p.1271. panel A (B) corresponds to the diversification measure $D_1 (D_2)$.

Choosing to underdiversify leads to a sacrifice in terms of variance to obtain a more positively skewed distribution of returns. However, it is interesting to assess whether this argument is the same when looking backward. In fact, when investors build their portfolios, their information is based on past data. Consequently, they may choose to underdiversify and select portfolios on past moments. Table 6 presents the same statistics as table 5 but the moments are estimated using the quarter preceding the portfolio formation date. The relationship between skewness in returns and diversification moves in the same direction; underdiversified portfolios exhibit a higher skewness, even if, in some quarters, the difference in skewness is not so large between deciles. Given our preceding theoretical and simulation results, these observations are not surprising.

[Insert Table 6 here]

[10] The results are similar for all periods and are available upon request.

When observing all quarters, skewness always decreases when diversification increases. However, over time, the mean level of skewness evolves according to market movements, and the same observation can be performed regarding the variation of skewness between the first and the last deciles. Figure 5 (resp. 6) represents the dynamics of median skewness for deciles 1 (resp. high diversification, identified by stars), 5 (resp. medium diversification-squares) and 10 (resp. low diversification-circles) for subsequent (resp. past) returns.

[**Insert Figure 5 here**]

[**Insert Figure 6 here**]

These graphs confirm that the difference of skewness between decile 10 (low diversification) and decile 1 (high diversification) is always positive but that this difference appears to be higher in bull market periods, namely, in the 5 first quarters and after quarter 20, that is, from the middle of 2003 to the end of 2006. This finding would be in complete agreement with our simulation results provided in table 2, if we could demonstrate that systematic variance is higher (as a percentage of total variance) in bear markets over our period of study (approximately between July 2000 up to April 2003). In the following subsection, we duplicate the methodology of section 2.2 and apply it to real stocks and investors to confirm this result.

4.2 Skewness, diversification and average correlation of returns

We decompose the period 1999-2006 in 32 quarters, with each of them being equivalent to a simulated market of section 2.2. There are approximately 65 trading days per quarter; each trading day is "equivalent" to a state of nature. Of course, the number of stocks held by at least one investor is significantly greater than the number of days in a quarter. Consequently, in addition to the general test with all stocks, we perform three other tests with different numbers of stocks. More precisely, we select stocks that are held by 0.5%, 1% and 1.5% of investors in each quarter. The aim of this selection process is to avoid "marginal stocks" held by a few investors and to prevent drawing general conclusions driven by the behavior of the least frequently traded stocks. Moreover, we also exclude the most illiquid stocks for which the proportion of zero returns is above 10% (Lesmond et al. 1999).

The results are summarized in table 7. The four panels are ranked according to the number of stocks taken into account. In each panel, we report the average number of stocks across quarters. For the constrained cases, the average number varies from 76 for stocks held by at least 1.5% of investors to 162 for stocks held by 0.5% of investors. We observe high and significant negative rank correlations in almost all cases, as in the simulation study. When all stocks are considered, the rank correlation varies between -0.724 in the one-factor model and -0.774 in the three-factor model. This variation is then extremely significant. However, even in the most constrained case with 76 stocks, the correlation is around -0.5, which is also significant at all usual levels.

This finding confirms the mechanical link between the decrease in skewness due to diversification and the variance due to common factors. A more synthetic statistic can be obtained by calculating for each quarter the average correlation coefficient of stock returns and then to calculate the rank correlation between the vector of average correlations and the vectors of decrease in skewness. This rank correlation varies from -0.31 to -0.59 across the four subsets of stocks. In other words, a strong link exists between skewness variation due to diversification choices and market conditions.

[**Insert Table 7 here**]

To complete our analysis, it is worth noting that the way in which we select stocks to produce table 7 may be questioned. In fact, choosing stocks held by a minimum percentage of investors biases the selection process towards diversified investors. In other words, doing so could eliminate stocks that are prominently held by skewness seekers holding very underdiversified portfolios. In short, our conclusion would be true but the behavioral interpretations of this conclusion could be flawed. A way to address this problem is to duplicate the results of table 7 with randomly chosen stocks. For panels A to C of table 7, we know the number of stocks in each quarter. Let n_i the number of stocks in quarter i for the panel under consideration. We draw stocks at random $(n_1, ..., n_{32})$ and perform the same calculations as in table 7. We repeat 100 times the sequence, evaluating the significance of the rank correlation in each simulation. Table 8 summarizes the results. Each panel provides the number of cases where the rank correlation is significant (it is always negative) at the 1%, 5% and 10% levels. For example, considering the first factor in panel A, 70 draws out of 100 give a significant correlation at the 1% level, 93 at the 5% level and, finally, 96 at the 10% level. Comparable results are obtained for panels B and C. It turns out that the decrease in skewness is more linked to diversification itself than to stock-picking skills by underdiversified investors.

[Insert Table 8 here]

Using a different methodology, Mitton and Vorkink (2007) found that investors holding underdiversified portfolios also pick highly skewed stocks. The above result does not support this conclusion. Consequently, hereafter, we duplicate the methodology of Mitton and Vorkink to evaluate our preceding result.

Mitton and Vorkink define a stock-specific statistic named SAID (security's average investor diversification), defined as follows:

$$SAID_k^l = \frac{1}{n_k} \sum\nolimits_{j=1}^{n_k} D_j^l \tag{13}$$

where D_j^l denotes the *l-th* diversification measure of investor j. n_k is the number of investors holding stock k. $SAID_k^l$ is then the average diversification measure of stock-k holders. If undiversified investors concentrate on highly skewed stocks, we should obtain a cross-sectional positive relationship between SAID and the skewness of returns. Highly skewed stocks should attract undiversified investors and increase the SAID of these stocks. Mitton and Vorkink (2007) obtain such a relationship by sorting stocks with respect to the variable SAID and by averaging skewness on single stocks within deciles. When adopting the same approach, we do not obtain such a relationship in all periods. To analyze this point more precisely, we calculate at the end of each year the cross-sectional rank correlation between SAID and skewness. If high SAID is linked to high skewness, the correlation should be significantly positive.

The results are reported in table 9, with the moments being calculated with one year of daily returns. The results are divided into four panels according to the diversification index (D_1 or D_2) and the period over which the returns are considered (past or subsequent).

[Insert Table 9 here]

For example, in panel A, the line identified by the year 1999 illustrates the number of stocks held by investors at the end of this year (column N), and ρ is the Spearman rank correlation between $SAID$ and the level of skewness over the set of $N = 1042$ stocks. The next columns contain the p-value of ρ, and the median levels of skewness in deciles 1, 5 and 10.

We focus our comment on panels C and D related to D_2 because it takes into account the portfolio weights. There is a clear break in 2003, that is at the market reversal when the market volatility began to decrease (and the idiosyncratic volatility started to increase). In the first part of the period, there is a significant positive relationship between SAID and D_2 meaning that people are concentrating on highly skewed stocks. Following this, no significant link appears between the two variables, which shows that we must be cautious before concluding that a part of investors are gamblers, not only chasing skewness by underdiversification but also selecting highly skewed stocks. This conclusion is not supported by our data and is clearly dependent on the evolution of the market.

4.3 Portfolio value and diversification choices

Table 3 illustrates the strong link between diversification and portfolio value. Market imperfections may be the source of the relationship between these two variables. If a per trade transaction cost is borne by individual investors, it becomes very costly to manage a "diversified" portfolio of stocks when the portfolio value is low. There is a strong incentive to focus on a low number of stocks. According to Liu (2010), it is also optimal for "poor" investors to underdiversify if they are subject to solvency constraints. Liu shows that below a given level of wealth, the optimal portfolio contains only one stock. The optimal number of stocks then increases with wealth. Consequently, solvency constraints can also explain the use of underdiversification among less wealthy investors.

However table 6 shows that less diversified portfolios generate higher skewness in returns. This phenomenon also appeared in our theoretical results in section 2.

To analyze more deeply the link between diversification and skewness seeking, it is necessary to control for other possible reasons to underdiversify, especially transaction costs and solvency constraints. The most natural means of doing so is to regress diversification measures on skewness, controlling for portfolio values in a panel data setting with fixed effects[11]. The model we use is the following:

$$D_{jt} = a_j + a_S Skewness_{jt} + a_P \ln(Portfolio\_Value_{jt}) + \varepsilon_{jt} \tag{14}$$

where $Skewness_{jt}$ is the skewness of past returns on the portfolio of investor j and $Portfolio\_Value_{jt}$ is the portfolio value of investor j at the date D_{jt} is measured, that is in the end of each quarter, semester or year, depending on the periodicity of the analysis.

Panel data allow us to control for unobservable variables such as personality traits. In other words, our model accounts for individual heterogeneity. Moreover, by using fixed effects, we explore the relationship between skewness, portfolio value and diversification within individual investors. Each investor has her own individual characteristics that may impact skewness, portfolio value or diversification (for instance, an investor's gender may influence the number of assets in a portfolio). When using fixed effects, we remove the effect of these time-invariant characteristics from the predictor variables and assess the net effect of predictors.

In table 10, we present quarterly, semi-annual and annual results over the period 1999-2006. For example, for semi-annual results, return skewness is estimated over 6 months of daily returns, and the panel data analysis is conducted over 16 semesters. For each period of time, the first (second) column refers to D_1 (D_2) as the diversification index. *N obs* and *N groups* refer to the number of rows and the number of individual investors (or individual portfolios). As our panel is unbalanced, individual investors were present in 18.9 quarters, 9.6 semesters and 4.98 years on average. In all models, the results are significant at the highest level. Three different measures of the reliability of our results are provided: the F-statistic (testing whether the vector of regression

[11] We significantly reject the hypothesis of a random effect model (Hausman χ^2 statistic is equal to 3358.10).

coefficients is the null vector), and the R^2s, overall, between and within, and finally the intraclass correlation rho, that is the fraction of the variance that is due to differences across individuals.

The first three lines provide the regression coefficients of equation 14 with t-stats between parentheses. For example, consider index D_1 in the semi-annual analysis. There are 77253 portfolios, $F = 33895.41$, the $\overline{R}^2$ coefficients are equal to 0.3848 (within), 0.5339 (between) and 0.4847 (overall). The intercept is 1.6728, the coefficient of $Portfolio\_Value$ is -0.1480 and the coefficient of portfolio skewness is 0.0102.

The main comments are identical for the three analyses. First, all three coefficients are significant at the 1% level. The most significant variable is clearly $Portfolio\_Value$ which has t-stats varying from -162.25 to -287. According to table 3, the sign of the coefficient and the statistical significance of this variable are not surprising. This result is in line with Liu's model (2010) as less wealthy investors subject to solvency constraints hold a portfolio containing a very low number of stocks. This point was not considered by Mitton and Vorkink (2007). Moreover, it appears that the coefficient of $Portfolio\_Value$ has higher t-stats for D_1 than for D_2. Therefore, the importance of portfolio value in determining the diversification level is higher when only the number of stocks is considered to measure diversification. This point is a supplementary argument in favor of the *fixed transaction cost* and *solvency constraint* approach.

However, what is also important is that skewness is always significant, for either D_1 or D_2, and that the regression coefficient is positive. Finally, the coefficient of $Skewness$ has higher t-stats for D_2 than for D_1. Considering different time-windows to estimate skewness does not change the results, as seen in table **10.** The regression coefficients are relatively stable across time-windows.

[Insert Table 10 here]

This stability means that, even after controlling for portfolio value, skewness remains significant (with the expected sign) in explaining diversification choices.

Finally, table 11 gives panel results of the estimation of equation (14) with indicator functions for semesters[12]. As coefficients for a_j, a_P and a_S remain stable, we note that the coefficients of

[12] The results over quarters and years yield the ame conclusions and are available upon request.

semester dummies reflect market trends. More precisely, all coefficients are negative but higher in absolute value in bear markets. The interpretation is that portfolio diversification is decreasing in bear markets, after controlling for portfolio value (which also decreases in bear markets) and skewness of returns.

5 Concluding remarks

Building highly skewed portfolios may reveal either a propensity to gamble or prudent behavior. In this paper, we first show analytically that diversification in Arrow-Debreu markets decreases the variance but also the skewness of returns. We also show that this relationship between diversification and skewness remains true on average in markets where the primary investments are portfolios of Arrow-Debreu securities.

We then analyze the behavior of a large sample of French individual investors and confirm some of the empirical results obtained by Mitton and Vorkink (2007) on U.S. investors. A lack of diversification provides higher skewness to investors at the expense of a higher variance of returns. This result is consistent with our theoretical results. However, when assessing whether undiversified investors concentrate on highly skewed stocks, we obtain unclear results. More precisely, the Spearman rank correlation between diversification measures and the skewness of stock returns does not present evidence of a persistent and significant relationship between these two variables. To dig deeper into this relationship, we link the decrease in skewness due to diversification to the share of variance explained by a given number of common factors. It turns out that the decrease in skewness is a decreasing function of this share of variance. In other words, in bearish markets characterized by a strong market factor and a high average correlation between stock returns, underdiversification does not significantly improve the level of skewness in portfolio returns.

Finally, we test whether skewness remains significant in the explanation of diversification choices after controlling for portfolio value. In fact, when investors do not invest much in the stock market, managing a diversified portfolio is costly, and investors may find it optimal to focus on a few stocks. Our panel data regression shows that this is the case; that is, portfolio value is the main determinant of diversification choices. Nevertheless, return skewness remains significant after controlling for portfolio value and time (market) effects.

Our contribution in this paper goes beyond that of Mitton and Vorkink in several directions. We first establish a theoretical link between diversification and skewness and then show that this relationship depends on the "concentration" of the market, that is, the sharing of global variance between idiosyncratic and systematic variance. We were able to analyze this aspect because of our new database of individual investors, which covers an eight-year period including the bursting of the dotcom bubble and the subsequent partial recovery.

References

Albuquerque, R. (2011), Skewness in Stock Returns: Reconciling the Evidence on Firm versus Aggregate Returns, Working Paper, Boston University.

Ang, A., Bekaert, G. (2002), International asset allocation with regime switching, *Review of Financial Studies*, 15(4), 1137-1187

Ang, A., Bekaert G. (2004), How do regimes affect asset allocation, *Financial Analysts Journal*, 60, 86-99

Baker, M., Wurgler, J. and Yuan, Y. (2012), Global, Local, and Contagious Investor Sentiment, *Journal of Financial Economics*, 104, 272-287.

Bali, T. G., Cakici, N., Whitelaw, R. F. (2011), Maxing out: Stocks as Lotteries and the Cross-Section of Expected Returns, *Journal of Financial Economics,* 99(2), 427-446.

Barberis, N., Huang, M. (2008), Stocks as Lotteries: the Implications of Probability Weighting for Security Prices. *American Economic Review*, 98, 2066–2100.

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.

Brunnermeier, M.K., Parker, J.A. (2005), Optimal Expectations, *American Economic Review,* 95, 1092-1118.

Brunnermeier, M.K., Gollier, C., Parker, J.A. (2007), Optimal Beliefs, Asset Prices and the Preference for Skewed Returns. *American Economic Review,* 97, 159–165.

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.

Campbell, R., Koedijk K., Kofman, P. (2002), Increased Correlation in Bear Markets, *Financial Analysts Journal*, 58, 87-94

Chamberlain, G., Rotschild, M. (1983), Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica*, 51(5), 1281-1304.

Chen J., Hong, H., Stein J., C. (2001), Forecasting crashes: Trading Volume, Past Returns, and Conditional Skewness in Stock Prices, *Journal of Financial Economics,* 61, 345–381.

Conine, Thomas E. Jr., and Maurry J. Tamarkin, 1981, On diversification given asymmetry in returns, *Journal of Finance,* 36, 1143-1155.

Cook, P. J., Clotfelter, C. T. (1993), The Peculiar Scale Economies of Lotto, *American Economic Review,* 83, 634–643.

Deck, C., Schlesinger, H. (2010), Exploring Higher-Order Effects, *Review of Economic Studies*, 77, 1403-1420.

Ebert, S., Wiesen, D. (2011), Testing for Prudence and Skewness Seeking, *Management Science,* 57(7), 1334-1349.

Eeckhoudt, L., Schlesinger, H. (2006), Putting Risk in its Proper Place, *American Economic Review,* 96(1), 280–289.

Forrest, D., Simmons, R., Chesters, N. (2002), Buying a Dream: Alternative Models of the Demand for Lotto, *Economic Inquiry*, 40, 485-496.

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

Garrett, T. A., Sobel, R. S. (1999), Gamblers Favor Skewness not Risk: further Evidence from United States' Lottery Games, *Economic Letters,* 63, 85–90.

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

Gollier, C. (2005), Optimal Illusions and Decision under Risk, University of Toulouse, IDEI Working Papers: No. 2376.

Harding, M. C. (2007), Explaining the Single Factor Bias of Arbitrage Pricing Models in Finite Samples, *Economics Letters,* 99, 85-88.

Harvey, C. R., Siddique, A. (1999), Autoregressive conditional skewness, *Journal of Financial and Quantitative Analysis,* 34, 465–487.

Harvey, C., Siddique, A. (2000), Conditional Skewness in Asset Pricing Tests, *Journal of Finance*, 55, 1263–1295.

Karolyi, A., Stulz, R. (2003), Are Financial Assets Priced Locally or Globally?, In: Constantinides, G., Harris,M., Stulz,R. (Eds.), The Handbook of the Economics and Finance.North-Holland Publishers, NewYork, pp. 975–1020.

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.

Kimball, M. (1990), Precautionary Saving in the Small and in the Large, *Econometrica,* 58(1), 53–73.

Kraus, A., Litzenberger, R.H. (1976), Skewness Preference and the Valuation of Risk Assets. *Journal of Finance,* 31, 1085–1100.

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

Kumar, A. (2009), Who Gambles in the Stock Market?, *Journal of Finance,* 64, 1889–1933

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

Lesmond, D., Ogden J., Trzcinka C. (1999), A new estimate of transaction costs, *Review of Financial Studies* 5,1113–1141.

Lewis, K. (1999), Explaining Home Bias in Equities and Consumption. *Journal of Economic Literature*, 37,571–608.

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

Longin, F., and Solnik, B. (2001), Correlation Structure of International Equity Market During Extremely Volatile Periods, *Journal of Finance*, 56(2), 649-676.

Markowitz, H.M. (1952), Portfolio Selection, *Journal of Finance*, 7, 77–91.

Menezes, C., Geiss, C., Tressler, J. (1980), Increasing Downside Risk, *American Economic Review,* 70(5), 921–932.

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

Nagel, S. (2005), Short Sales, Institutional Investors and the Cross-Section of Stock Returns, *Journal of Financial Economics*, 78, 277-309.

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

Shefrin, H., Statman, M. (2000), Behavioral Portfolio Theory, *Journal of Financial and Quan-*

titative Analysis, 35 (2), 127-151.

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

Tarazona-Gomez, M. (2004), Are Individuals Prudent? An Experimental Approach Using Lottery Choices, Working Paper, Copenhagen Business School.

Tversky, A., Kahneman, D. (1992), Advances in Prospect Theory : Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty*, 5, 297–323.

Walker, Y., Young, J. (2001), An Economist's Guide to Lottery Design, *The Economic Journal,* 111, 700-722.

Xu, J. (2007), Price Convexity and Skewness, *Journal of Finance*, 62(5), 2521-2552.

APPENDIX

Proof. of proposition 1 The first point is obvious since states are equally-likely. $\sigma_i^2 = E(X_i^2) - E(X_i)^2 = \frac{1}{n} - \frac{1}{n^2}$ since $X_i^m = X_i$ for any positive integer m. The third central moment is calculated as follows

$$E\left[(X_i - \mu_i)^3\right] = E(X_i^3) - 3\mu_i E(X_i^2) + 3\mu_i^2 E(X_i) - \mu_i^3 \tag{15}$$

$$= \frac{1}{n} - 3\frac{1}{n^2} + 3\frac{1}{n^3} - \frac{1}{n^3} = \frac{1}{n} - \frac{3}{n^2} + \frac{2}{n^3} \tag{16}$$

$$= \frac{(n-1)(n-2)}{n^3} \tag{17}$$

Finally, we get $cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = -1/n^2$ since $X_i X_j \equiv 0$ when $i \neq j$. ■

Proof. of proposition 2 Proposition **1** allows to write the covariance matrix of the n AD securities payoffs as

$$\mathbf{V}_n = \frac{1}{n}\mathbf{I}_n - \frac{1}{n^2}\mathbf{1}_{(n,n)} \tag{18}$$

where $\mathbf{I}_n$ is the (n, n) identity matrix and $\mathbf{1}_{(n,n)}$ is a (n, n) matrix containing only ones. As $p_k = \frac{1}{k}\sum_{i=1}^{k} X_i$, we get

$$\sigma_k^2 = \frac{1}{k^2}\mathbf{1}'_{(k)}\mathbf{V}_k\mathbf{1}_{(k)}$$

where $\mathbf{1}_{(k)}$ denotes a column vector of ones with k components and $\mathbf{V}_k$ the square matrix of the first k rows and columns of $\mathbf{V}_n$.

Equation (18) implies $\mathbf{V}_k = \frac{1}{n}\mathbf{I}_k - \frac{1}{n^2}\mathbf{1}_{(k,k)}$. We then write

$$\sigma_k^2 = \frac{1}{k^2}\mathbf{1}'_{(k)}\mathbf{V}_k\mathbf{1}_{(k)} = \frac{1}{k^2}\mathbf{1}'_{(k)}\left(\frac{1}{n}\mathbf{I}_k - \frac{1}{n^2}\mathbf{1}_{(k,k)}\right)\mathbf{1}_{(k)} \tag{19}$$

$$= \frac{1}{k^2 n}\mathbf{1}'_{(k)}\mathbf{I}_k\mathbf{1}_{(k)} - \frac{1}{k^2 n^2}\mathbf{1}'_{(k)}\mathbf{1}_{(k,k)}\mathbf{1}_{(k)} \tag{20}$$

$$= \frac{1}{kn} - \frac{1}{n^2} = \frac{1}{n}\left(\frac{1}{k} - \frac{1}{n}\right) \tag{21}$$

As expected, the variance of the equally-weighted portfolio decreases with the number of AD securities in the portfolio. The case $k = n$ gives $\sigma_n^2 = 0$ which is consistent withe fact that p_n is a risk-free portfolio paying $1/n$ in each state. ■

Proof. of proposition 3

$$s_k^3 = \frac{1}{k^3} E\left[(Y_k^3 - \left(\frac{k}{n}\right)^3 - 3\left(\frac{k}{n}\right) Y_k^2 + 3\left(\frac{k}{n}\right)^2 Y_k \right] \tag{22}$$

$$= \frac{1}{k^3}\left[\frac{k}{n} - \left(\frac{k}{n}\right)^3 - 3\left(\frac{k}{n}\right)^2 + 3\left(\frac{k}{n}\right)^3 \right] \tag{23}$$

Rearranging terms leads to

$$s_k^3 = \frac{1}{n^3}\left[\left(\frac{n}{k}\right)^2 - 3\left(\frac{n}{k}\right) + 2 \right] \tag{24}$$

$$= \frac{1}{n^3}\left[\left(\frac{n}{k} - 1\right)\left(\frac{n}{k} - 2\right) \right] \tag{25}$$

■

Table 1

Second and third moments as a function of the number of Arrow-Debreu securities in the portfolio

The first column gives the number of Arrow-Debreu securities in portfolios, columns 2 and 3 provide the variance and third central moments of portfolio payoffs. Columns 4 to 6 are defined as columns1 to 3 for portfolios containing 11 to 19 Arrow-Debreu securities.

Number of AD securities	σ_k^2 ($\times 10^3$)	s_k^3 ($\times 10^4$)	Number of AD securities	σ_k^2 ($\times 10^3$)	s_k^3 ($\times 10^4$)
1	47.500	42.750	11	2.045	-0.019
2	22.500	9.000	12	1.667	-0.028
3	14.167	3.306	13	1.346	-0.031
4	10.000	1.500	14	1.071	-0.031
5	7.500	0.750	15	0.833	-0.028
6	5.833	0.389	16	0.625	-0.023
7	4.643	0.199	17	0.441	-0.018
8	3.750	0.094	18	0.278	-0.012
9	3.056	0.034	19	0.132	-0.006
10	2.500	0.000			

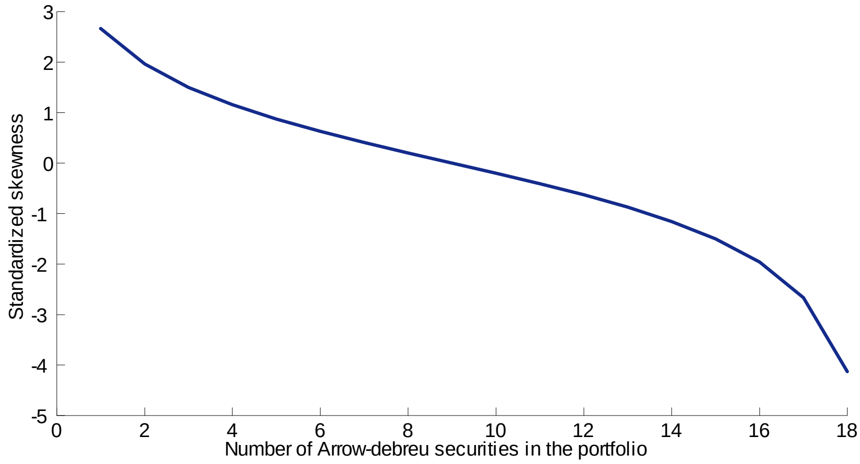

Figure 1

Evolution of standardized skewness as a function of the number of AD securities in the portfolio. In this example, $n = 20$.

Table 2

Correlation between skewness decrease due to diversification and cumulated variance of common factors

The table provides the cumulated percentage of variance for k factors and the Spearman rank correlation between the decrease in skewness due to diversification and the sum of the k first eigenvalues for k=1,..,5. Correlations are calculated for 1,000 simulated markets. Panel A (B, C) corresponds to markets containing 20 (60, 100) assets with 20 (60, 100) states of nature. The variation of skewness is calculated as the difference between the average skewness of single stocks and the skewness of the equally-weighted market portfolio.

Number of factors (k)	1	2	3	4	5
Panel A: 1000 simulated markets, 20 assets					
Cumulated variance (in %)	24.06	41.19	54.15	64.15	72.06
Rank Correlation	-0.469	-0.443	-0.391	-0.317	-0.252
Panel B: 1000 simulated markets, 60 assets					
Cumulated variance (in %)	14.27	24.79	33.14	40.04	45.88
Rank Correlation	-0.496	-0.519	-0.488	-0.448	-0.414
Panel B: 1000 simulated markets, 100 assets					
Cumulated variance (in %)	11.79	20.52	27.51	33.32	38.25
Rank Correlation	-0.525	-0.568	-0.552	-0.515	-0.479

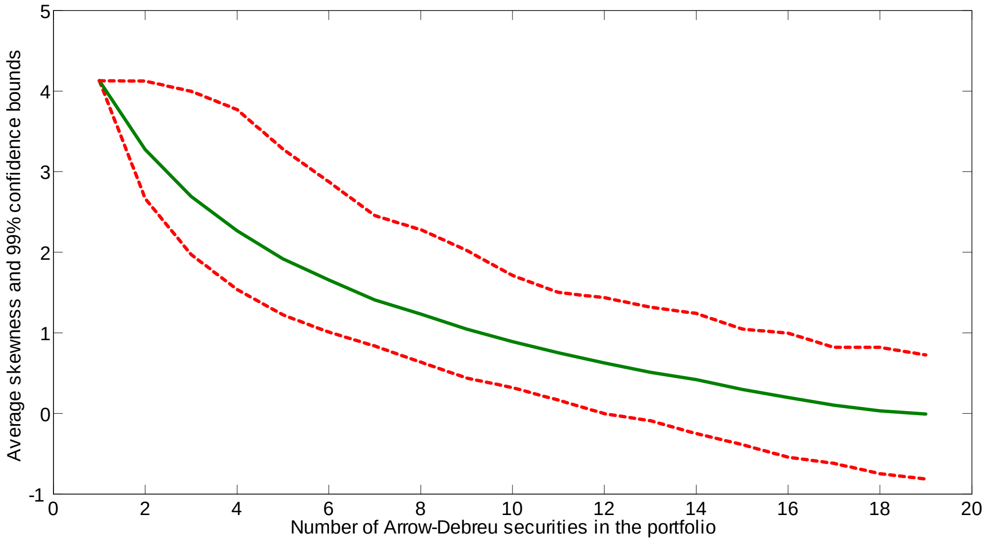

Figure 2

The solid line gives the evolution of the average skewness of portfolios as a function of the number of Arrow-Debreu securities in portfolios. Each point is the average over 1,000 simulated markets. The dashed lines represent the corresponding 99% confidence bounds, that is the 5th lowest and highest skewness obtained over 1,000 simulations.

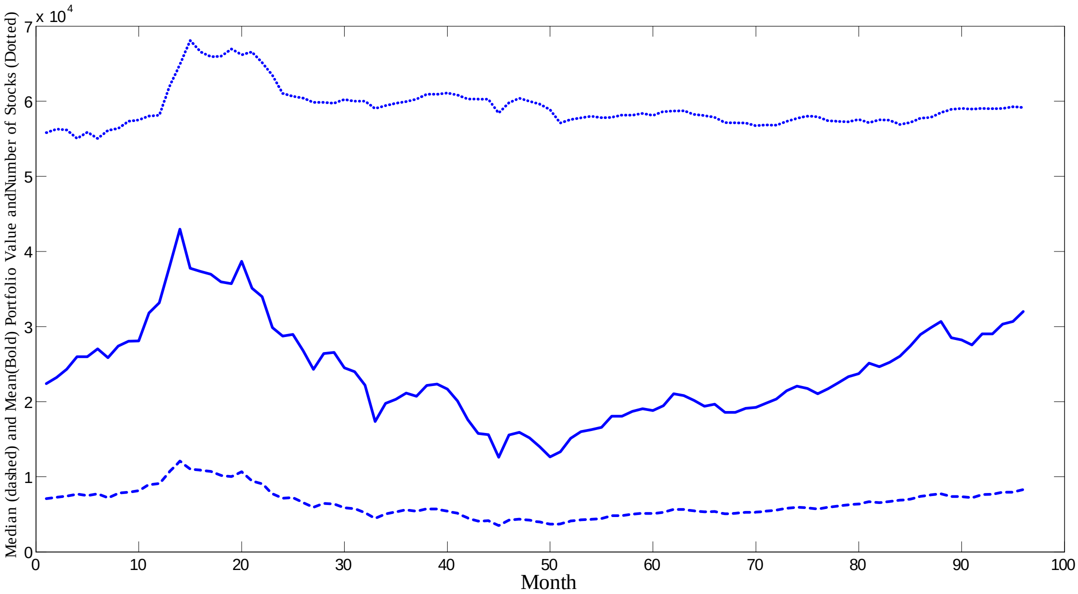

Figure 3

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.

Table 3

Descriptive statistics about portfolio values at three points in time

The table gives descriptive statistics about portfolios held by investors at three points in time, July 2000 (Panel A), July 2003 (Panel B) and July 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.

Number of Stocks	Number of Investors	Mean Port. Value	1st quartile	Median Port. Value	3rd quartile
					Panel A: Portfolios as of July 2000
1	8,956	6,856.56	662.10	1,571.00	3,979.34
2	7,007	9,538.46	1,778.26	3,453.00	7,497.21
3	5,900	15,786.24	3,206.64	5,817.43	11,784.08
4	4,731	18,606.83	4,708.57	8,430.00	16,186.30
5	3,704	23,432.00	6,399.34	11,004.70	20,061.54
6 to 9	9,688	37,683.76	10,097.43	17,240.40	32,473.37
More than 10	10,791	98,816.47	25,711.11	46,660.96	87,891.66
All	50,777	35,992.91	3,430.11	10,254.33	28,730.22
					Panel B: Portfolios as of July 2003
1	13,197	2,535.42	250.20	603.20	1,591.80
2	8,927	4,695.01	774.53	1,646.20	3,596.42
3	6,815	7,207.04	1,444.63	2,815.98	5,973.14
4	5,320	8,368.42	2,167.75	4,178.13	8,376.70
5	4,097	12,819.20	3,099.98	5,697.86	11,416.66
6 to 9	10,123	18,788.08	5,371.32	9,649.33	18,374.48
More than 10	10,307	50,762.58	14,457.45	26,714.47	50,355.61
All	58,786	15,903.96	1,248.30	4,241.70	13,318.64
					Panel B: Portfolios as of July 2006
1	10,426	4,295.99	390.40	1,003.00	2,740.20
2	6,913	8,637.35	1,279.61	2,815.50	6,401.13
3	5,240	12,643.92	2,443.90	5,106.74	11,053.84
4	3,987	17,353.15	3,943.36	7,437.04	16,041.48
5	3,158	22,100.53	5,327.29	9,986.42	19,985.22
6 to 9	7,678	33,021.51	9,095.00	16,809.07	33,416.07
More than 10	8,096	88,715.78	24,546.38	46,556.67	92,137.80
All	45,498	28,166.44	2,038.15	7,324.47	23,801.08

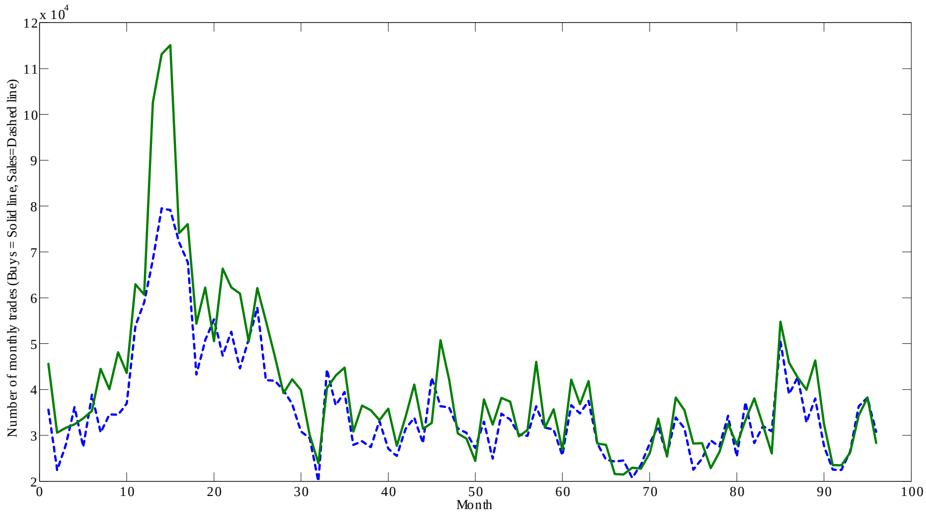

Figure 4

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

Table 4

Trades and holdings for stocks of the 6 main countries

Panel A indicates how trades are shared among the six most active countries of origin of traded stocks. Panel B gives the number of investors in the database at the end of each year and the percentage of investors holding stocks of the six countries coded as follows: FR = France, NL = The Netherlands, US = United States, GB = United Kingdom, DE = Germany, IT = Italy. Percentages in a given line sum above 1 due to international diversification of some investors.

Panel A: Trades in stocks of the 6 main countries							
	Total	FR	NL	US	GB	DE	IT
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 5

Moments of subsequent returns of investors' portfolios sorted by diversification levels

This table provides the return statistics of portfolios held by investors over the quarter following portfolio formation, Moments of returns are averaged within each decile of diversification (D_1 on the left of the table and D_2 on the right). Column N provides the number of investors in deciles. Return (R) and standard deviation (σ) are annualized. All moments, including skewness (S_k) are estimated on one quarter of daily data

	Subsequent returns sorted on D_1					Subsequent returns sorted on D_2				
July 00	N	D_1	R	σ	S_k	N	D_2	R	σ	S_k
H Div.	4,055	0.044	-0.071	0.186	0.072	4,560	0.074	-0.052	0.174	0.096
2	4,003	0.074	-0.085	0.205	0.095	4,560	0.125	-0.081	0.202	0.126
3	4,535	0.102	-0.100	0.220	0.114	4,560	0.170	-0.093	0.221	0.134
4	4,622	0.135	-0.109	0.234	0.120	4,560	0.218	-0.111	0.237	0.140
5	3,194	0.167	-0.118	0.247	0.139	4,560	0.273	-0.117	0.254	0.162
6	3,640	0.200	-0.111	0.255	0.153	4,560	0.342	-0.121	0.273	0.166
7	4,643	0.250	-0.123	0.274	0.182	4,566	0.431	-0.165	0.297	0.163
8	5,730	0.333	-0.141	0.291	0.176	4,554	0.528	-0.149	0.316	0.189
9	6,618	0.500	-0.170	0.329	0.214	4,560	0.739	-0.259	0.381	0.141
L Div.	7,509	1.000	-0.218	0.401	0.255	7,509	1.000	-0.218	0.401	0.255
July 03	N	D_1	R	σ	S_k	N	D_2	R	σ	S_k
H Div.	4,124	0.046	0.436	0.187	0.317	4,951	0.081	0.454	0.185	0.313
2	4,621	0.080	0.428	0.202	0.310	4,952	0.137	0.438	0.204	0.308
3	5,631	0.114	0.419	0.214	0.316	4,951	0.186	0.433	0.217	0.304
4	2,659	0.143	0.419	0.224	0.324	4,952	0.239	0.429	0.228	0.318
5	3,250	0.167	0.408	0.229	0.324	4,951	0.297	0.403	0.236	0.339
6	4,032	0.200	0.417	0.238	0.343	4,952	0.364	0.409	0.252	0.359
7	5,164	0.250	0.402	0.248	0.354	4,951	0.457	0.376	0.260	0.373
8	6,612	0.333	0.407	0.263	0.382	4,952	0.539	0.383	0.277	0.410
9	8,470	0.500	0.380	0.288	0.429	4,951	0.728	0.316	0.286	0.470
L Div.	11,403	1.000	0.331	0.330	0.536	11,403	1.000	0.331	0.330	0.536
July 06	N	D_1	R	σ	S_k	N	D_2	R	σ	S_k
H Div.	3,343	0.046	0.151	0.148	-0.092	3,776	0.074	0.150	0.147	-0.100
2	3,541	0.080	0.164	0.158	-0.033	3,775	0.129	0.163	0.158	-0.052
3	2,516	0.106	0.162	0.164	-0.021	3,776	0.179	0.160	0.167	-0.033
4	3,723	0.135	0.164	0.171	-0.015	3,775	0.234	0.156	0.175	-0.002
5	2,389	0.167	0.165	0.175	0.032	3,776	0.297	0.164	0.184	0.018
6	3,083	0.200	0.158	0.183	0.026	3,775	0.368	0.156	0.193	0.039
7	3,862	0.250	0.160	0.190	0.059	3,776	0.468	0.149	0.202	0.070
8	5,036	0.333	0.148	0.202	0.060	3,775	0.553	0.138	0.213	0.100
9	6,487	0.500	0.133	0.221	0.120	3,776	0.765	0.157	0.227	0.205
L Div.	8,815	1.000	0.084	0.264	0.178	8,815	1.000	0.084	0.264	0.178

Table 6

This table provides the return statistics of portfolios held by investors over the quarter preceding portfolio formation, Moments of returns are averaged within each decile of diversification (D_1 on the left of the table and D_2 on the right). Column N provides the number of investors in deciles. Return (R) and standard deviation (σ) are annualized. All moments, including skewness (S_k) are estimated on one quarter of daily data

	Past returns sorted on D_1					Past returns sorted on D_2				
July 00	N	D_1	R	σ	S_k	N	D_2	R	σ	S_k
H Div.	4,033	0.044	0.057	0.263	-0.086	4,383	0.073	0.066	0.239	-0.113
2	3,976	0.074	0.067	0.280	-0.046	4,384	0.123	0.065	0.272	-0.033
3	4,486	0.102	0.058	0.299	-0.006	4,383	0.167	0.049	0.291	0.001
4	4,516	0.135	0.050	0.315	0.030	4,384	0.214	0.043	0.310	0.035
5	3,102	0.167	0.033	0.328	0.051	4,383	0.267	0.028	0.332	0.062
6	3,553	0.200	0.036	0.341	0.077	4,384	0.335	0.026	0.356	0.081
7	4,412	0.250	-0.006	0.364	0.088	4,383	0.420	-0.009	0.388	0.089
8	5,394	0.333	0.002	0.384	0.113	4,384	0.526	-0.007	0.395	0.144
9	5,979	0.500	-0.034	0.423	0.154	4,383	0.738	-0.063	0.496	0.174
L Div.	6,666	1.000	0.002	0.484	0.205	6,666	1.000	0.002	0.484	0.205
July 03	N	D_1	R	σ	S_k	N	D_2	R	σ	S_k
H Div.	4,131	0.046	0.978	0.232	-0.033	5,006	0.081	0.957	0.230	-0.040
2	4,644	0.080	0.964	0.252	0.023	5,007	0.138	0.966	0.256	0.021
3	5,660	0.114	0.958	0.266	0.053	5,006	0.187	0.967	0.270	0.058
4	2,672	0.143	0.944	0.279	0.083	5,007	0.240	0.956	0.286	0.108
5	3,268	0.167	0.932	0.285	0.111	5,006	0.299	0.946	0.297	0.134
6	4,067	0.200	0.949	0.296	0.140	5,007	0.366	0.956	0.319	0.178
7	5,250	0.250	0.931	0.311	0.161	5,006	0.459	0.940	0.332	0.228
8	6,717	0.333	0.930	0.333	0.212	5,007	0.540	0.910	0.357	0.239
9	8,649	0.500	0.891	0.369	0.279	5,006	0.729	0.856	0.370	0.294
Low Div.	11,585	1.000	0.755	0.413	0.343	11,585	1.000	0.755	0.413	0.343
July 06	N	D_1	R	σ	S_k	N	D_2	R	σ	S_k
H Div.	3,372	0.046	-0.171	0.213	-0.376	3,816	0.074	-0.175	0.214	-0.383
2	3,571	0.080	-0.160	0.218	-0.336	3,816	0.129	-0.171	0.221	-0.340
3	2,542	0.106	-0.169	0.223	-0.312	3,816	0.179	-0.180	0.224	-0.307
4	3,762	0.135	-0.176	0.227	-0.295	3,816	0.234	-0.178	0.230	-0.276
5	2,421	0.167	-0.172	0.229	-0.275	3,817	0.297	-0.179	0.236	-0.258
6	3,120	0.200	-0.169	0.234	-0.246	3,816	0.369	-0.179	0.242	-0.215
7	3,897	0.250	-0.172	0.241	-0.213	3,816	0.470	-0.181	0.249	-0.184
8	5,073	0.333	-0.184	0.250	-0.183	3,816	0.555	-0.182	0.258	-0.157
9	6,587	0.500	-0.176	0.263	-0.127	3,816	0.767	-0.106	0.268	-0.080
L Div.	9,100	1.000	-0.084	0.295	-0.019	9,100	1.000	-0.084	0.295	-0.019

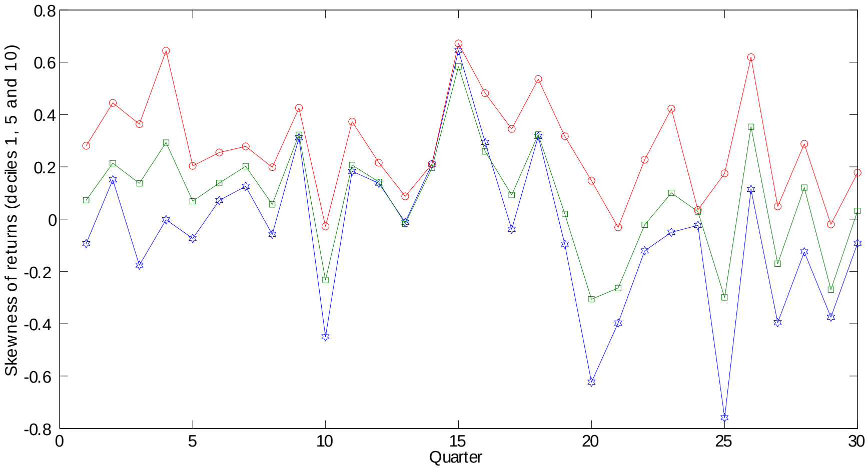

Figure 5

Evolution of median portfolio skewness over the period 1999-2006 for deciles 1 (high diversification identified by stars), 5 (squares) and 10 (low diversification, circles) for subsequent returns

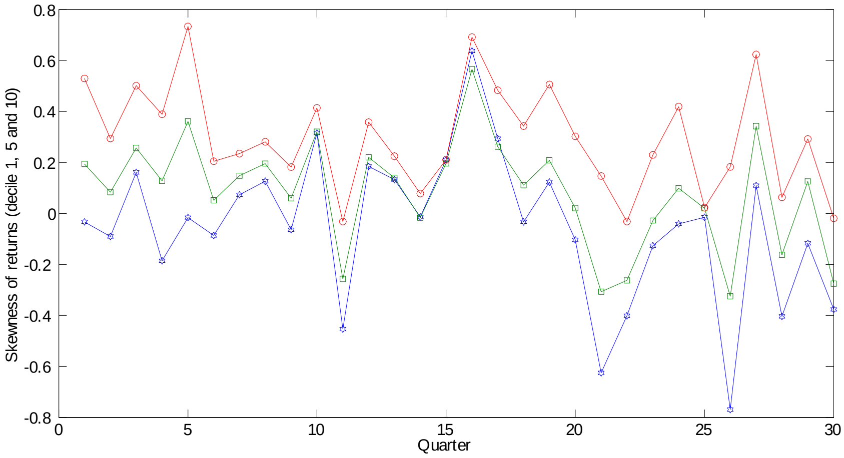

Figure 6

Evolution of median portfolio skewness over the period 1999-2006 for deciles 1 (high diversification identified by stars), 5 (squares) and 10 (low diversification, circles) for past returns

Table 7

Rank correlations between the decrease in skewness due to diversification and the share of variance on the common factors

The table provides the cumulated percentage of variance for k factors and the Spearman rank correlation between the variation of skewness and the sum of the k first eigenvalues for k=1,..,5. Correlations are calculated for different subsets of stocks defined according to a minimum percentage (1.5, 1, 0.5 for Panels A to C) of investors holding these stocks. Panel D is based on the complete set of stocks in each quarter, that is stocks held by at least one investor. The variation of skewness is calculated as the difference between the average skewness of single stocks and the skewness of the equally-weighted market portfolio.

Number of factors (k)	1	2	3	4	5
Panel A: 76 stocks (average) held by at least 1.5% of investors					
Cumulated variance (in %)	31.78	40.22	46.02	50.72	54.79
Rank Correlation	-0.567^{***}	-0.484^{***}	$-0;494^{***}$	-0.466^{***}	-0.468^{***}
Panel B: 102 stocks (average) held by at least 1% of investors					
Cumulated variance (in %)	28.54	36.52	42.14	46.70	50.66
Rank Correlation	-0.636^{***}	-0.546^{***}	-0.481^{***}	-0.465^{***}	-0.441^{**}
Panel C: 162 stocks (average) held by at least 0.5% of investors					
Cumulated variance (in %)	23.71	30.29	35.50	39.86	43.57
Rank Correlation	-0.533^{***}	-0.611^{***}	-0.593^{***}	-0.540^{***}	-0.521^{***}
Panel D: all stocks					
Cumulated variance (in %)	17.37	23.17	27.23	30.54	33.55
Rank Correlation	-0.724^{***}	-0.75^{***}	-0.774^{***}	-0.768^{***}	-0.762^{***}

Table 8

Percentages of significant rank correlations between the decrease in skewness due to diversification and the share of variance on the common factors

The table contains three panels corresponding to the numbers of stocks of the three panels of the preceding table. In each panel, we give the percentage of draws leading to a significant rank correlation (at the 1, 5 and 10 percent levels) between the decrease in skewness and the share of variance in the first 5 factors. The variation of skewness is calculated as the difference between the average skewness of single stocks and the skewness of the equally-weighted market portfolio.

Number of factors (k)	1	2	3	4	5
Panel A: 76 stocks (average) randomly chosen					
Percentage of significant rank correlations at the 1% level	70	57	52	45	44
Percentage of significant rank correlations at the 5% level	93	85	76	76	76
Percentage of significant rank correlations at the 10% level	96	91	88	87	86
Panel B: 102 stocks (average) randomly chosen					
Percentage of significant rank correlations at the 1% level	78	70	69	69	68
Percentage of significant rank correlations at the 5% level	95	88	88	89	88
Percentage of significant rank correlations at the 10% level	99	95	95	93	93
Panel C: 162 stocks (average) randomly chosen					
Percentage of significant rank correlations at the 1% level	84	85	79	76	75
Percentage of significant rank correlations at the 5% level	97	97	97	91	90
Percentage of significant rank correlations at the 10% level	99	98	99	97	97

Table 9

Rank correlations between SAID of single stocks and skewness of returns

This table provides the Spearman rank correlation between the SAID and the skewness of one year of daily returns. N is the number of stocks held at the end of the corresponding year. *Rho* is the Spearman rank correlation coefficient between D and skewness. pval is the p-value of the correlation, The three last columns give the median skewness in deciles 1 (high diversification), 5 (medium diversification) and 10 (low diversification). Panels A and B refer to the diversification index D_1 where in A (B) past(subsequent) returns are considered to calculate skewness. Panels C and D are defined similarly but the diversification index is D_2.

Panel A: Past returns D_1

	N	ρ	pval	Hdiv	Mdiv	Ldiv
1999	1042	0.02	0.60	0.36	0.58	0.51
2000	1232	0.03	0.24	0.34	0.49	0.46
2001	1318	0.08	0.00	0.21	0.22	0.43
2002	1296	0.06	0.02	0.14	0.30	0.32
2003	1239	0.02	0.49	0.53	0.41	0.40
2004	1196	-0.03	0.29	0.41	0.44	0.29
2005	1182	0.02	0.51	0.52	0.43	0.31
2006	1143	0.01	0.63	0.34	0.52	0.40

Panel B: Subsequent returns D_1

	N	ρ	pval	Hdiv	Mdiv	Ldiv
1999	1141	0.06	0.04	0.34	0.61	0.46
2000	1315	0.11	0.00	0.18	0.28	0.45
2001	1315	0.08	0.00	0.17	0.20	0.39
2002	1221	0.04	0.20	0.52	0.39	0.40
2003	1189	-0.04	0.13	0.34	0.35	0.17
2004	1142	-0.01	0.69	0.58	0.41	0.53
2005	1165	0.01	0.86	0.52	0.50	0.39
2006	1169	0.01	0.78	0.43	0.37	0.26

Panel C: Past returns D_2

	N	ρ	pval	Hdiv	Mdiv	Ldiv
1999	1042	0.04	0.23	0.37	0.70	0.66
2000	1232	0.06	0.04	0.34	0.49	0.45
2001	1318	0.12	0.00	0.16	0.37	0.43
2002	1296	0.08	0.00	0.14	0.24	0.33
2003	1239	0.03	0.35	0.45	0.46	0.40
2004	1196	-0.05	0.07	0.44	0.44	0.15
2005	1182	0.01	0.64	0.50	0.49	0.31
2006	1143	0.04	0.21	0.32	0.53	0.40

Panel D: Subsequent returns D_2

	N	ρ	pval	Hdiv	Mdiv	Ldiv
1999	1141	0.08	0.01	0.32	0.45	0.47
2000	1315	0.13	0.00	0.27	0.45	0.41
2001	1315	0.11	0.00	0.16	0.25	0.35
2002	1221	0.06	0.05	0.47	0.35	0.42
2003	1189	-0.05	0.09	0.27	0.40	0.07
2004	1142	0.01	0.81	0.55	0.40	0.48
2005	1165	0.02	0.51	0.36	0.53	0.31
2006	1169	0.03	0.32	0.42	0.28	0.23

Table 10

Panel regression of diversification index on portfolio value and skewness of returns.

The six columns correspond to two diversification indices (D_1 and D_2) and three ways to calculate the skewness of returns (with a quarter, a semester, or a year of daily data). The regression equation is the following :

$$D_{jt} = a_j + a_S Skewness_{jt} + a_P ln(PortfolioValue_{jt}) + \varepsilon_{jt}$$

a_j denotes the intercept, a_P is the coefficient of portfolio value and a_S is the coefficient of skewness. *Nobs* is the number of observations, that is the sum of the numbers of investors over the periods. *Ngroups* the number of different investors entering the analysis. Two different investors may stay in the database for different time-lengths. R^2 (within, between, overall) are the determination coefficients of the regression. F is the Fisher statistic and *rho* denotes the intraclass correlation. As usual, *** denotes statistical significance at the 1% level.

	Quaterly		*Half − yearly*		*Yearly*	
	D_1	D_2	D_1	D_2	D_1	D_2
a_j	1.6694*** (372.96)	1.5633*** (306.71)	1.6728*** (332.70)	1.5610*** (271.69)	1.6339*** (271.82)	1.5380*** (227.58)
a_P	−.1474*** (−287.00)	−.1260*** (−215.48)	−.1480*** (−258.33)	−.1259*** (−192.45)	−.1440*** (−212.05)	−.1238*** (−162.25)
a_S	.0081*** (41.57)	.0111*** (54.75)	.0102*** (32.77)	.0148*** (45.52)	.0051*** (10.62)	.0102*** (20.16)
F	41793.16***	24745.08***	33895.41***	19887.69***	22636.72***	13738.01***
R^2 *within*	0.3820	0.2900	0.3848	0.2868	0.3732	0.2778
R^2 *between*	0.5452	0.4908	0.5339	0.4789	0.5050	0.4501
R^2 *overall*	0.4894	0.4130	0.4847	0.4086	0.4716	0.3990
rho	0.7068	0.6796	0.7173	0.6878	0.7248	0.6967
N obs.	1549842	1549842	739317	739317	339125	339125
N groups	82110	82110	77253	77253	68204	68204

Table 11

Panel regression of diversification index on portfolio value, skewness of returns and dummy variables for semesters.

The two left columns correspond to diversification index D_1 and the two right columns to D_2 (with associated t-statistics). The regression equation is the following:

$$D_{jt} = a_j + a_S Skewness_{jt} + a_P ln(PortfolioValue_{jt}) + a_1 S_1 + a_2 S_2 ..... + \varepsilon_{jt}$$

a_j denotes the intercept, a_P is the coefficient of portfolio value and a_S is the coefficient of skewness. The variables S_t are indicator functions for the semesters. $Nobs$ is the number of observations, that is the sum of the numbers of investors over the periods. $Ngroups$ the number of different investors entering the analysis. Two different investors may stay in the database for different time-lengths. R^2 (within, between, overall) are the determination coefficients of the regression. F is the Fisher statistic and rho denotes the intraclass correlation. As usual, *** denotes statistical significance at the 1% level.

	D_1	*t-stat*	D_2	*t-stat*
a_j	1.837***	330.63	1.705***	268.65
a_P	−0.160***	−259.39	−0.1357***	−193.23
a_S	0.0134***	41.15	0.0180***	52.54
$S1$	−0.0341***	−21.8	−0.0272***	−15.93
$S2$	−0.0086***	−5.85	−0.0123***	−7.62
$S3$	−0.0197***	−14.71	−0.0297***	−19.94
$S4$	−0.0448***	−36.12	−0.0526***	−38.33
$S5$	−0.0635***	−54.41	−0.0667***	−51.39
$S6$	−0.0753***	−67.03	−0.0788***	−64.7
$S7$	−0.0989***	−88.91	−0.0865***	−71.6
$S8$	−0.1277***	−112.92	−0.1168***	−96.91
$S9$	−0.1152***	−105.93	−0.1108***	−96.63
$S10$	−0.0891***	−86.76	−0.0891***	−82.99
$S11$	−0.0687***	−70.02	−0.0668***	−65.36
$S12$	−0.0603***	−63.82	−0.0592***	−60.00
$S13$	−0.0418***	−46.91	−0.040***	−43.66
$S14$	−0.0255***	−31.94	−0.0263***	−31.85
$S15$	−0.0158***	−24.31	−0.0135***	−20.20
F	4412.72		2752.93	
R^2 *within*	0.4279		0.3242	
R^2 *between*	0.5385		0.4821	
R^2 *overall*	0.4948		0.4178	
rho	0.733		0.699	
N obs.	739317		739317	
N groups	77253		77253	

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