# In search of positive skewness: the case of individual investors

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

June 2012

**Abstract**

We first prove analytically in an Arrow-Debreu market that the skewness of portfolio returns 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 more than 80,000 individual investors at a large brokerage house. Investors underdiversify their portfolios to get higher skewness in returns, a result in line with the one obtained by Mitton and Vorkink (2007) on a sample of US investors. Unlike these authors, we do not find a clear evidence that underdiversified investors concentrate on highly skewed stocks. However, we show that the decrease in skewness due to diversification is essentially driven by the share of total variance of stock returns explained by common factors.

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

---

[*]LARGE Research Center, EM Strasbourg Business School, University of Strasbourg, 61 avenue de la forêt noire, 67085 Strasbourg CEDEX, France. We thank Laurent Deville, Louis Eeckhoudt, Gunter Franke, Christophe Godlewski, 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 Meeting (2011, München) for comments and suggestions. The financial support of OEE (Observatoire de l'Epargne Européenne) is gratefully acknowledged. We thank Tristan Roger for valuable research assistance and computer programming. E-mails of authors: proger@unistra.fr, mhb@unistra.fr, merli@unistra.fr. Corresponding author: Patrick Roger, 0033368852156.

The standard theory of portfolio choice developed by Markowitz (1952) leads to the conclusion that investors should hold a diversified portfolio. In the mean-variance framework, investors are seeking returns with a high first moment and a low second moment. 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, Mitton and Vorkink 2007, Kumar 2007, Goetzman and Kumar 2008) also show on large samples of individual U.S investors that portfolios they hold are largely underdiversified, containing less than five stocks on average[1].

Several theoretical explanations have been given to underdiversification. Some are linked to skewness seeking by investors, others are based on the existence of solvency constraints (Liu 2010) and/or market imperfections (transaction costs for example).

A number of models show that investors may be skewness seekers if one of the following conditions is satisfied:

1) their preferences are driven by a three-moment model (Mitton and Vorkink 2007);

2) they obey prospect theory (Barberis and Huang 2008);

3) they obey optimal expectations theory (Brunnermeier and Parker, 2005, Brunnermeier *et al.*, 2007)

Full diversification may not be optimal for such investors because diversification, though reducing variance of returns, reduces skewness as well. In the next section, we focus on this point by first proving analytically in an Arrow-Debreu setting that skewness of portfolio returns decreases with the number of Arrow-Debreu securities in portfolios. 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 the decrease of skewness due to diversification remains

[1] Calvet *et al.* (2007) obtained the same results for Sweden except that Swedish investors seem to have a little bit more diversified portfolios than US investors.

true when portfolios are not equally-weighted.

On the one hand, to the question of why investors underdiversify, Mitton and Vorkink (2007) answer that it is a way to capture higher skewness in portfolio returns. On the other hand, Harvey and Siddique (1999, 2000) and Chen *et al.* (2001)[2] show that the average skewness of single stocks is positive in most periods when the market skewness is negative most of the time.

In the empirical part of the paper, we analyze the behavior of 87,373 retail investors over a eight-year period (1999-2006). First duplicating some of the analyses of Mitton-Vorkink on our database, we confirm their results linking negatively the skewness of investors' portfolio returns to the level of diversification. Unlike these authors, we do not find that underdiversified investors consistently pick highly skewed stocks. It leads us to show that the decrease in skewness due to diversification is essentially linked to the share of aggregate variance due to common factors (by opposition to idiosyncratic volatility). We base our analysis on the sequence of eigenvalues of the covariance matrix of returns. In particular, measuring the decrease in skewness due to diversification by the difference between average skewness of single stocks and skewness of largely diversified portfolios, we show there is a strong negative rank correlation between this decrease and the share of variance due to common factors (measured by the sum of the first eigenvalues). It reinforces our preceding result showing that underdiversification is the essential mean to capture high skewness.

The paper is organized as follows. The first section reviews the literature on individual portfolios' (under)diversification. In section 2, we present the analytical and simulation results in Arrow-Debreu markets. Section 3 provides our empirical results and the paper concludes by directions of future research.

[2]More recently, Albuquerque (2011) shows that the average skewness of single stocks (estimated with six months of daily data) is always positive except during the second half of 1987 (due to the Black Monday). Over the same period, the skewness of the equally-weighted market portfolio is negative 77% of the time.

# 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 prefer positively skewed portfolios in the hope of getting 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 2010). On the opposite, preferring high skewness may simply reveal a prudent behavior in the sense of Kimball (1990). In this case, investors are said downside risk averse (Menezes *et al.* 1980) and their utility function is mainly characterized by a positive third-order derivative. These investors want to avoid large losses and it may be taken into account by overweighting the probabilities of low outcomes. In short, a gambler mainly overweights the probability of very high returns and slightly rebalances her beliefs over a large range of returns. The downside risk averse investor exagerates the probability of large losses and may rebalance either on very high returns (if she is a gambler too, like state-lottery players) or on a large range of returns. In all cases, she prefers positively skewed portfolios but the desired profile of returns may be different for the two categories of investors.

Underdiversification may then be justified if investors choose their optimal portfolios not only by considering the first two moments of the distribution of returns, but also the third one. The first attempt to introduce the third moment in portfolio choice was proposed by Kraus and Litzenberger (1976), followed by Harvey and Siddique (2000) who provided empirical support to this model.

Several recent experimental studies also show that around 60% of participants are prudent, when faced to lottery choices. Their decision process is then compatible with positive skewness seeking (Tarazona-Gomes 2004, Ebert and Wiesen 2011, Deck and Schlesinger 2010). Seeking positive skewness and making prudent choices is generally interpreted in these studies as downside risk aversion (Menezes *et al.* 1980, Eeckhoudt and Schlesinger, 2006).

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 (undesirable) idiosyncratic volatility but also reduces (desirable) positive skewness.

In the most recent (non expected utility) models, desirability of positive skewness also appears because investors distort probabilities. Shefrin and Statman (2000), in their behavioral portfolio theory, consider that investors' decisions are guided 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).

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. However, the transformation of probabilities in Cumulative Prospect Theory is "exogenous" in the sense that it depends only on the ranking of outcomes, not on their values.

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 maximizes the average of her current utility and her expected future utility. But the current utility is higher when the investor is optimistic about future prospects (anticipatory utility). This distortion of beliefs leads to suboptimal investment decisions in terms of ressource 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 the one 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 kind of "jackpot effect", like 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). On 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 on gambling markets like state lotteries. More precisely those who are prone to bet on state lotteries are also prone to choose low-priced stocks with high idiosyncratic risk and high positive skewness. Bali, Cakici and Whitelaw (2010) do not analyze the behavior of individual investors but rank 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. 35% of stocks in the highest decile one month are in the same decile the following month.

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

## 2.1 Moments of portfolios of Arrow-debreu securities

In this section we establish 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 simply measured by the inverse of the number of stocks in portfolios, the variance of portfolio returns is an linear function of the diversification index and the third central moment of returns is a quadratic function of the diversification measure.

Consider a finite state-space $\Omega = \{\omega_1, ..., \omega_n\}$ with $n$ equally-likely states of nature and assume that all Arrow-Debreu securities, denoted $X_1, ...X_n$, are traded, where $X_i$ pays 1 in state $\omega_i$ and 0 elsewhere. We consider a sequence $(p_1, ..., p_n)$ of equally-weighted portfolios respectively containing 1, 2...$n$, AD securities. Without loss of generality, we assume that $p_i$ contains $1/i$ units of each of the first $i$ securities.

We first recall elementary properties of the moments of AD securities before analyzing portfolios.

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

$$E(X_i) = m_i = \frac{1}{n} \quad (1)$$

$$V(X_i) = \frac{n-1}{n^2} \quad (2)$$

$$E\left[(X_i - m_i)^3\right] = \frac{(n-1)(n-2)}{n^3} \quad (3)$$

$$cov(X_i, X_j) = -\frac{1}{n^2} \text{ if } i \neq j \quad (4)$$

**Proof.** See the Appendix ■

Consider now a portfolio $p_k$ invested in the first $k$ AD securities and denote $\mu_k$ and $\sigma_k^2$ the expectation and the variance of payoffs.

**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 as a measure of diversification (denoted $D_1$ by Mitton and Vorkink (2007)). Proposition 2 shows that the variance of returns increases linearly in $D_1$. When $k = n$, the portfolio is risk-free and the variance is equal to 0.

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

$$s_k^3 = E((\frac{1}{k}\sum\nolimits_{i=1}^{k} X_i - \mu_k)^3) = \frac{1}{k^3}E((\sum\nolimits_{i=1}^{k} X_i - \frac{k}{n})^3) \tag{6}$$

Defining $Y_k = \sum\nolimits_{i=1}^{k} X_i$, $s_k^3$ is equal to $\frac{1}{k^3}E((Y_k - \frac{k}{n})^3)$. 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 $X_i^m X_j^t = 0$ for any pair $(m, t)$ of strictly positive integers and different indices $i$ 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 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 get that the third order moment increases quadratically in $D_1$. In fact, we have:

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

We can also establish a very simple relationship between $s_k^3$ and $\sigma_k^2$ 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 faster. It implies that standardized skewness, defined as $s_k^3/\sigma_k^3$ 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 (most of the time) is often explained by overreaction to good news and underreaction to bad news (Nagel, 2005, Xu, 2007). When estimating skewness of a single-stock with a time-series of returns, it is likely 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 do not come together for all stocks in the portfolio. On the contrary, it is well documented that correlations of stock returns increase in hard times. Consequently, low returns are more likely to be observed simultaneously, leading to negative skewness for portfolio returns. A "similar" shape is observed with portfolio of Arrow-Debreu securities which pay 1 on different states of nature.

**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 to get simple analytical results about the evolution of skewness as a function of the number of Arrow-Debreu securities. To test our result in a more general framework, we simulate portfolio weights in the following way. Still considering 20 states of nature, we randomly select 1000 portfolios for each number of securities between 2 and 19. For a $m$-security portfolio, we draw $m$ random numbers $x_1, ..., x_m$ between 0 and 100 and define the weights as $w_i = x_i / \sum_{j=1}^{m} x_j$. We consider only positive weights since individual investors almost never use short selling.

**Insert Figure 2 around here**

Figure 2 shows the evolution of the average skewness (bold line) and the corresponding 99% confidence bounds. Of course, there is no uncertainty for single-stock portfolios because of the analytical solution of proposition 1. We observe that the average skewness is always positive. It comes close to 0 only for the maximum number of stocks. It could seem surprising at a first glance since, in the equally-weighted case of the preceding section, the skewness becomes negative when $k > n/2$. The reason is that portfolio weights are simulated according to a uniform multidimensional distribution. It turns out that the proportion of portfolios close to the equally-weighted case is very low and these are the portfolios with the lowest skewness. Comparing the bold curve on figure 2 to the results obtained by Mitton and Vorkink (2007, table 3, p1271) shows a similar evolution of skewness with respect to diversification. Consequently, it appears that underdiversification is a good way to capture skewness. It is even not necessary to try picking highly skewed stocks to get a highly skewed portfolio.

### 2.2.2 The general case

The assumption of Arrow-Debreu securities traded on the market is rather simplistic. To generalize the approach, we consider now $n$ stocks traded on the $n$-state market, each stock being itself a portfolio of AD securities. To generate the payoffs of stock $i$, we first draw at random a number $n_i$ of AD-securities in the set of $n$ assets. We then define the payoffs of stock $i$ by drawing $n_i$ random weights (summing to one), as in the preceding section. We therefore build a $(n, n)$ matrix of payoffs of single stocks. It corresponds to one simulated market.

In most simulations, the dimension of the space spanned by the stocks is equal to $n$. However, these payoffs are more or less correlated, depending on random draws. Our aim is at illustrating the link between portfolio skewness and average correlations. More precisely, empirical studies show that skewness decreases with diversification, but the steepness of the slope changes over time. For example, 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.

In section 2.2.1 we saw that the average skewness is a decreasing convex function of the number of AD securities in portfolios. In other terms, the average skewness is a decreasing function of the dimension of the space spanned by the subset of AD securities in portfolios. The generalization of this idea in a financial market context is as follows. If stock returns are highly correlated, the share of variance coming from common factors in stock returns is high. We then expect a slow decrease in skewness due to diversification simply because the efficiency of diversification is lower. In fact, increasing the number of stocks in a portfolio does not increase proportionately the dimension of the space spanned by the stocks included in the portfolio.

On the opposite, when the average correlation of stock returns is low (corresponding to high idiosyncratic volatilities), the decrease in skewness due to diversification should be larger.

In a single(market)-factor model, the share of market variance is equal to the percentage of the total variance lying in the first eigenvalue of the covariance matrix of returns. More generally, in a $k$-factor model, the variance due to common factors is the sum of the $k$ first eigenvalues. Our conjecture is then that the average skewness decreases faster with diversification when the variance due to common factors is low, in percentage of total variance. 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 a principal component analysis (as suggested by Chamberlain and Rotschild, 1983) is difficult since 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 $k = 5$ 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 $k$ first factors where $k = 1, .., 5$. The second line provides the Spearman rank correlation of this percentage of variance with the decrease in skewness. Of course, these correlations being valued over 1,000 simulations, they are significant at all the standard significance levels.

In fact, it is worth mentioning that a low percentage of simulations leads to an increase in skewness (6.6%, 2.4%, 2.7% respectively in panels A, B and C), meaning that the equally-weighted portfolio is more skewed than the average stock. These cases correspond to high levels of market variance since 30.1% (20.03%, 16.8%) of variance is on the first factor, compared to an average of 24.06% (14.27%, 11.79%) in table 2.

These first results show that if underdiversification generates skewness, a given level of diversification is likely to provide different levels of skewness depending on market conditions. In particular, it is well known that bearish markets tend to increase the average correlation between stock returns, meaning that the percentage of variance lying in the common factors is higher. We then observe a lower skewness of returns. We come back to this point in the next section devoted to the empirical study of a large sample of individual investors.

# 3 Data and descriptive statistics

## 3.1 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 nine million trades realized by 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 3 shows three time-series. The upper dotted curve represents the average number of stocks held by investors ($\times 10^4$). It 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 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 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. 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 respectively 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 average portfolio values in our sample are not different 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 more than one million euros.

Table 3 gives some more detailed statistics at three points in time, January 2000, January 2003 and January 2006[3]. We use the same presentation as 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 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 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 in the two kinds of trades up to April 2000. Around 110,000 monthly buy trades were realized

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

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.

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 4 around here**]

## 3.2 Stock data

Price data come from two sources, Eurofidai for stocks traded on Euronext[4] 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 investment 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 4. It illustrates the well-known home bias puzzle[5]. It is 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 on 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. 96.97% held French stocks (meaning that

[4]http://www.eurofidai.org. A part of this database has been recently used by Foucault *et al.* (2011) in their study of retail trading and volatility on the French market and by Baker *et al.* (2010) to study the contagion of sentiment across countries, including France and the U.S.

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

around 3% held only foreign stocks), 21.05% were holding Dutch stocks but only 3.97% were holding 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 (on real markets) whether underdiversified portfolios should bear more skewness. To answer this question we first 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[6]. We first use returns subsequent to the portfolio formation date since the realized skewness for a given portfolio is the skewness of future returns. However, on a behavioral point of view, investors can base their choices on the observation of past skewness. We then check if 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.

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

Though simple, this measure does not take into account the weighting of securities within portfolios. Consequently, we also 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 returns. One advantage of equation (12) is that it is standardized by variance (or standard deviation). It is a way to take into account the mechanical positive link between variance and skewness 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 first analyze the link between underdiversification and skewness in returns. In a second part, we perform a stock-level analysis to answer the following question: do underdiversified investors concentrate on highly skewed stocks? We do not get a clearcut answer to this question, the strength of the relationship essentially depends on the direction and volatility of the market. To confirm this conjecture, we finally link the decrease in skewness due to diversification with the average percentage of variance in returns due to common factors.

## 4.1 The investor-level analysis

### 4.1.1 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 in our sample period. As we study portfolio skewness, we first calculate returns on these portfolios for the subsequent quarter. We then calculate the average value of $\widehat{S_k}^3$ within each decile (as well as $\overline{r}$ and $\widehat{\sigma}$ which are annualized in table 5) . We cannot present here the results for all quarters so we take the same three points in time as before, but the results are similar on all periods.

[**Insert Table 5 here**]

Table 5 provides the results when subsequent returns are used to estimate moments. There is a stable relationship between the diversification level and the second and third moments of the distribution of returns. This relationship is increasing and our results are in line with those of Mitton and Vorkink (2007). We adopt a presentation similar to their table 3 p.1271. Panel A (B) corresponds to the diversification measure $D_1(D_2)$. When looking at all quarters, the skewness always decreases when diversification increases. However, over time the mean level of skewness evolves according to market movements.

Choosing to underdiversify corresponds to a sacrifice in terms of variance to obtain a more positively skewed distribution. However, it is interesting to know if the 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. On past data, the relationship between skewness in returns and diversification goes in the same direction, underdiversified portfolios exhibit a higher skewness even if, in some quarters, the difference in skewness is not large between deciles.

[**Insert Table 6 here**]

To give an idea of what happens on the entire period, we represent the dynamics of median skewness on deciles 1 (high diversification-stars), 5 (medium diversification-squares) and 10 (low diversification-circles) in figure 5 for subsequent returns and on figure 6 for past returns.

[**Insert Figure 5 here**]
[**Insert Figure 6 here**]

This graph confirms that the difference of skewness between decile 10 (low diversification) and decile 1 (low diversification) is always positive. This difference is 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. Meanwhile, the average level of skewness largely decreases and is negative on some quarters. It is important to note that we analyze investors portfolios, some of them experiencing large losses after the internet bubble burst. Consequently, it may happen that some of them leave the market (and the database) after losses. It is then impossible to conclude that some investors capture a persistent positive skewness in returns. The other remark is that an investor-level analysis is quite different from the stock-level analysis performed in the next section, simply because some stocks are held by a very low number of investors and other by a very large number of investors. As a consequence, the two analyses answer two completely different questions.

#### 4.1.2 Controlling for portfolio value

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 certain wealth level, the optimal portfolio contains only one stock. The optimal number of stocks then increases with wealth. Consequently, solvency constraints can also justify underdiversification for the less wealthy investors.

But 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 way to do that is to regress diversification measures on skewness, controlling for portfolio values in a panel data setting with fixed effects[7]. The model we use is the following:

$$D_{jt} = a_j + a_S Skewness_{jt} + a_P \ln(Portfolio\_Value_{jt}) + \varepsilon_{jt} \tag{13}$$

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.

Panel data allows us to control for variables we cannot observe or measure like personal factors. In other words, our model accounts for individual heterogeneity. Moreover, by using fixed-effects we explore the relationship between skewness and portfolio's values and diversi-

[7] We significantly reject the hypothesis of a random effect model (Hausman Chi2 statistic is 3358.10).

fication within individual investors. Each investor has his own individual characteristics that may impact skewness, portfolios' values or diversification (for instance, being a male or a female may influence the number of assets in portfolio). When using fixed-effects, we remove the effect of these time-invariant characteristics from the predictor variables and assess the predictors' net effect.

In table 7, we present quarterly, semi-annual and annual results over 1999-2006. For example, for semi-annual results, return skewness is estimated over 6 months of daily data 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, on the average individual investors were present 18.9 quarters, 9.6 semesters and 4.98 years. In all models, results are significant at the highest level. More accurately, we obtain 3 measures of the reliability of our results: the $F$-statistic (test whether all the coefficients in the model are different than zero), the $R^2$ overall, between and within and the intraclass correlation $rho$, that is the fraction of the variance that is due to differences across individuals.The first three lines provide the coefficients $a_j$, $a_P$ and $a_S$ of the variables, the $t-$statistics and the level of significance. For example, consider index $D_1$ with half-yearly periods. There are 77253 portfolios, $F = 33895.41$, the $\overline{R}^2$ are equal to 0.3848, 0.5339 and 0.4847, the intercept is 1.6728, the coefficient of $Portfolio\_Value$ is $-0.1480$ and the coefficient of the portfolio skewness is 0.0102;all coefficients are significant at the 1% level.

The main comments are identical for all periods. The most significant variable is the $Portfolio\_Value$ with $t$-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) since "poor" 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 $D_2$. It means that the importance of portfolio value in determining the diversification level is higher when only the number of stocks is considered to measure diversification. It is an argument in favor of the "fixed transaction cost" approach.

But what is also important is the fact that skewness is always significant for $D_1$ and $D_2$,with a positive coefficient. Moreover, it appears that the coefficient of $Skewness$ has higher $t$-stats for $D_2$ than for $D_1$.

[**Insert Table 7 here**]

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

## 4.2 The stock-level analysis

It appears at the investor-level that investors choosing to underdiversify their portfolios are looking for positive skewness. The question is then to know if they not only select a low number of stocks but pick stocks whose returns are highly positively skewed. In fact, if underdiversified investors are "pure gamblers" they should choose to concentrate their holdings on highly skewed stocks. If they underdiversify for other reasons, for example prudence, it may not be the case.

To decide which assumption is the most likely, Mitton and Vorkink define a stock specific statistic they name SAID (security's average investor diversification) defined as follows:

$$SAID_k^j = \frac{1}{n_k} \sum\nolimits_{i=1}^{n_k} D_i^j \tag{14}$$

where $D_i^j$ denotes the $j-th$ diversification measure of investor $i$. $n_k$ is the number of investors holding stock $k$. $SAID_k^j$ is then the average diversification measure of stock-$k$ holders. If undiversified investors concentrate on highly skewed stocks, we should obtain a positive relationship between the SAID and the skewness of stock returns in the cross section. 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 methodology, 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 of returns. The results are reported in tables 8 and 9, the moments being calculated with one year of daily returns. Consequently, the statistics provided in the different lines of the tables are obtained over non overlapping periods.

[**Insert Table 8 here**]

[**Insert Table 9 here**]

Table 8 is related to the diversification measure $D_1$, that is the inverse of the number of stocks. Panel A(B) provides the results when skewness is based on past (subsequent) returns. For example, in panel A, the line identified by the year 1999 gives the number of stocks held by investors at the end of this year (column N), the Spearman rank correlation between $D_1$ and the level of skewness of the set of $N = 1042$ stocks, the $p$-value of the correlation and the median levels of skewness in decile 1, 5 and 10, named Hdiv, Mdiv and Ldiv. Table 9 provides the same information with the diversification measure $D_2$, that is the Herfindahl index of the portfolio weights. We only comment here this second table because $D_2$ takes into account the portfolio weights. There is a clear break in 2003, that is at the market reversal when the market volatility started 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 people are concentrating on highly skewed stocks. After that, no significant link appears between the two variables. This shows that we need to be cautious before concluding that a part of investors are pure gamblers, not only chasing skewness by underdiversification but also selecting highly skewed stocks.

## 4.3 Skewness, diversification and average correlation of returns

In this section we duplicate the methodology of section 2.2 and apply it to real stocks and investors. We decompose the period 1999-2006 in 32 quarters, each of them being equivalent to a simulated market of the preceding section. The number of trading days is around 65 per quarter; it is identified to the number of states of nature. Of course, the number of stocks held by at least one investor is largely 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 respectively 0.5%, 1% and 1.5% of investors in each quarter under consideration. 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 most unfrequently traded stocks.

The results are summarized in table 10. The four panels are ranked according to the number of stocks taken into account. In each panel, the number of stocks is the average number across quarters. It varies from 81 for stocks held by at least 1.5% of investors to 664 in the case without constraints. We observe high and significantly negative rank correlations in almost all cases. It confirms the results obtained for simulations in the preceding section and illustrates 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 table 10.

**[Insert Table 10 here]**

## 4.4 Robustness checks

The way we select stocks in table 10 may be questioned for the following reason. Choosing stocks held by a minimum percentage of investors biases the selection process towards diversified investors. In other words, we 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 deal with this problem is to duplicate the results of table 10 with randomly chosen stocks. For panels A to C of table 10, 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 at random $(n_1, ...., n_{32})$ stocks and perform the same calculations as in table 10. We repeat 100 times the sequence. Table 11 summarizes the significance results. A given Panel gives the number of cases where the rank correlation is significant (they are 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 11 here**]

# 5 Concluding remarks

Building highly skewed portfolios may reveal either a propensity to gamble or a prudent behavior. In this paper we first show analytically that diversification in Arrow-Debreu markets decreases variance and skewness of returns. We also show that this relationship between diversification and skewness remains true on average in markets where primary securities are portfolios of Arrow-Debreu securities. We then analyze the behavior of a large sample of French individual investors. We confirm some of the results already obtained by Mitton and Vorkink (2007) on U.S investors. Lack of diversification provides higher skewness to investors at the expense of a higher variance of returns. This result is consistent with the theoretical approach. However, when looking whether undiversified investors concentrate on highly skewed stocks, we get unclear results. More precisely, calculating the Spearman rank correlation between diversification measures and skewness of stock returns does not show evidence of a persistent and significant relationship between these two variables. To dig deeper in 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 much improve the level of skewness in portfolio returns.

# References


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

[2] 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.

[3] Barber, B. M., Odean, T. (2000), Trading is hazardous to your wealth: The Common Stock Investment Performance of Individual Investors, *Journal of Finance,* 55, 773–806.

[4] Barber, B. M., Odean, T. (2001), Boys will be Boys: Gender, Overconfidence, and Common Stock Investment, *Quarterly Journal of Economics* 116, 261–292.

[5] Barber, B., Odean, T., and M. Strahilevitz, M. (2004), Once Burned, Twice Shy: Naïve Learning, Counterfactuals, and the Repurchase of Stocks Previously Sold, working paper,University of California.

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

[7] 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.

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

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

[10] 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.

[11] 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.

[12] 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.

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

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

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

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

[17] Farrell, L., Morgenroth, E., Walker, I. (1999) A Time Series Analysis of UK Lottery Sales: the Long Run Price Elasticity, *Oxford Bulletin of Economics and Statistics*, 61, 513–26.

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

[19] Foucault, T., Sraer D., Thesmar, D. (2011), Individual Investors and Volatility, *Journal of Finance*, forthcoming.

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

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

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

[23] Graham, J., Kumar A. (2006), Do Dividend Clienteles Exist? Evidence on Dividend Preferences of Retail Investors, *Journal of Finance*, 61(3), 1305–1336.

[24] Grote, K.R., Matheson, V.A. (2007), Examining the 'Halo Effect' in Lotto Games, *Applied Economics Letters*, 2007, 14, 307–310

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

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

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

[28] Ivkovic, Z., Weisbenner, S. (2005), Local Does As Local Is: Information Content of the Geography of Individual Investors' Common Stock Investments, *Journal of Finance*, 60, 267–306.

[29] 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.

[30] 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.

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

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

[33] Kraus, A. & Litzenberger, R. H. (1983), On the Distributional Conditions for a Consumption-oriented three Moment CAPM, *Journal of Finance,* 38, 1381–1391.

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

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

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

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

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

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

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

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

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

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

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

[45] Shefrin, H., Statman, M. (2000), Behavioral Portfolio Theory, *Journal of Financial and Quantitative Analysis*, 35 (2), 127-151.

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

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

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

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

[50] Wang, J. H., Tzengb, L. Y., Tienc, J. (2006), Willingness to Pay and the Demand for Lotto, *Applied Economics*, 38, 1207-1216.

[51] 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

$$\begin{aligned} E\left[(X_i - \mu_i)^3\right] &= E(X_i^3) - 3\mu_i E(X_i^2) + 3\mu_i^2 E(X_i) - \mu_i^3 && (15)\\ &= \frac{1}{n} - 3\frac{1}{n^2} + 3\frac{1}{n^3} - \frac{1}{n^3} = \frac{1}{n} - \frac{3}{n^2} + \frac{2}{n^3} && (16)\\ &= \frac{(n-1)(n-2)}{n^3} && (17)\end{aligned}$$

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

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

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

$$\begin{aligned} \sigma_k^2 &= \frac{1}{k^2}\mathbf{1}'_{(k)}\mathbf{V}_k\mathbf{1}_{(k)} = \frac{1}{k^2}\mathbf{1}'_{(k)}\left(\frac{1}{n}\mathbf{I}_k - \frac{1}{n^2}\mathbf{1}_{(k,k)}\right)\mathbf{1}_{(k)} && (19)\\ &= \frac{1}{k^2 n}\mathbf{1}'_{(k)}\mathbf{I}_k\mathbf{1}_{(k)} - \frac{1}{k^2 n^2}\mathbf{1}'_{(k)}\mathbf{1}_{(k,k)}\mathbf{1}_{(k)} && (20)\\ &= \frac{1}{kn} - \frac{1}{n^2} = \frac{1}{n}\left(\frac{1}{k} - \frac{1}{n}\right) && (21)\end{aligned}$$

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**

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

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

Rearranging terms leads to

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

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

■

**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 columns 1 to 3 for 11 to 19 Arrow-Debreu securities.

| Number of AD securities | $\sigma_k^2$ $(\times 10^3)$ | $s_k^3$ $(\times 10^4)$ | Number of AD securities | $\sigma_k^2$ $(\times 10^3)$ | $s_k^3$ $(\times 10^4)$ |
|---|---|---|---|---|---|
| 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 | | | |

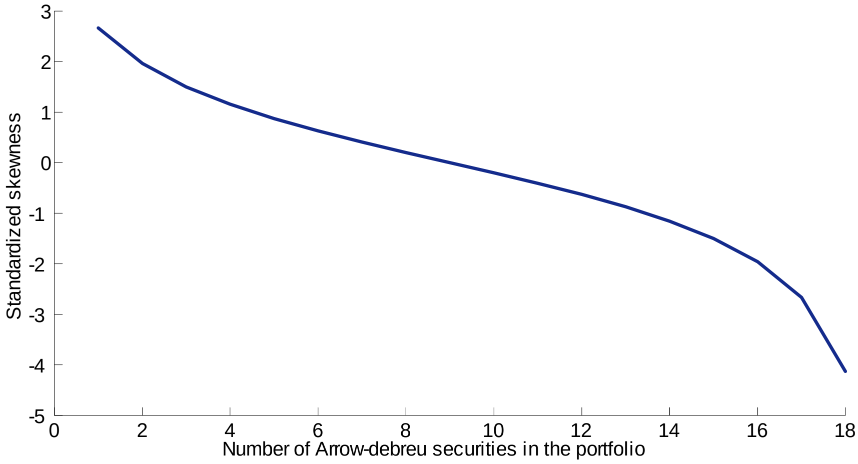


**Figure 1**
Evolution of standardized skewness as a function of the number of AD securities in the portfolio. In this example, $n = 20$.

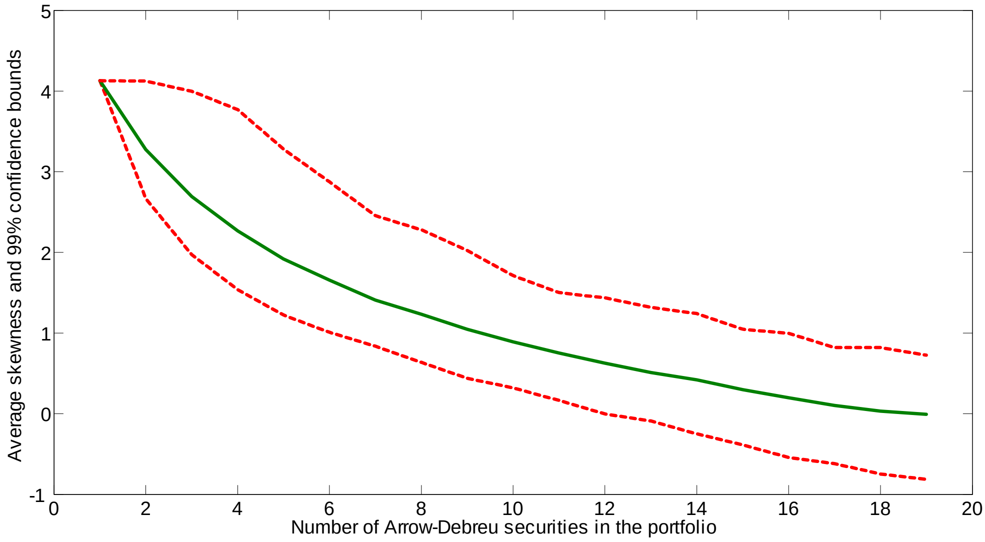


**Figure 2**
Average skewness and 99% confidence bounds as a function of the number of Arrow-Debreu securities in the portfolio.

**Table 2**
**Correlation between skewness variation and cumulated variance of 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 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 |

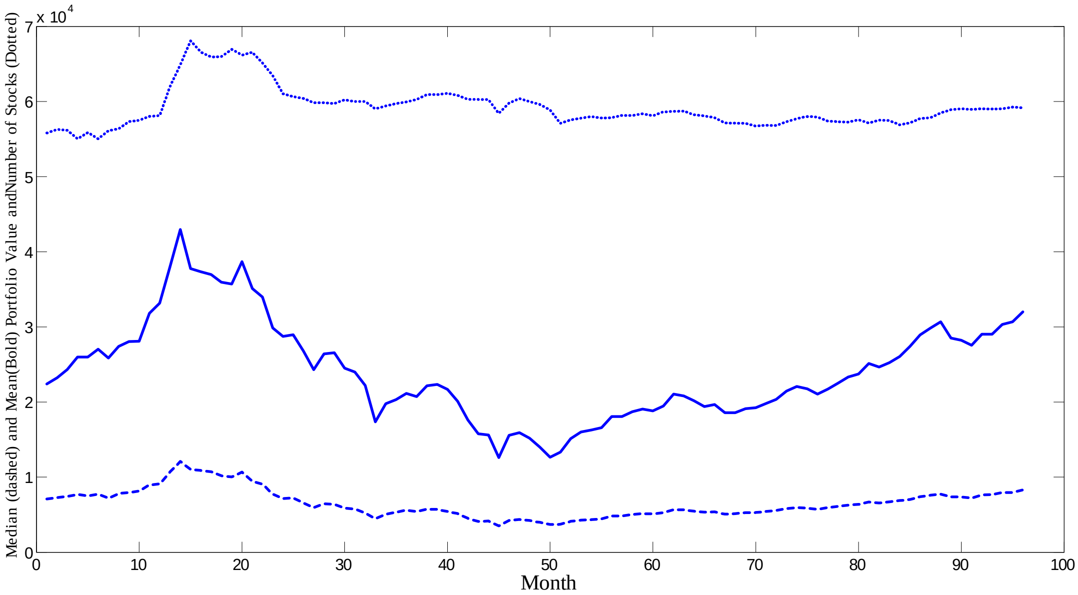


**Figure 3**
The three curves represent respectively the time-series of the average number of stocks held by investors, and the mean and median portfolio value. The period under consideration starts in January 1999 (month 1) and ends in December 2006 (month 96). The upper dotted curve is the average 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**
Statistics on portfolio values 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 | 8956 | 6856,56 | 662,10 | 1571,00 | 3979,34 |
| 2 | 7007 | 9538,46 | 1778,26 | 3453,00 | 7497,21 |
| 3 | 5900 | 15786,24 | 3206,64 | 5817,43 | 11784,08 |
| 4 | 4731 | 18606,83 | 4708,57 | 8430,00 | 16186,30 |
| 5 | 3704 | 23432,00 | 6399,34 | 11004,70 | 20061,54 |
| 6 to 9 | 9688 | 37683,76 | 10097,43 | 17240,40 | 32473,37 |
| More than 10 | 10791 | 98816,47 | 25711,11 | 46660,96 | 87891,66 |
| **All** | 50777 | 35992,91 | 3430,11 | 10254,33 | 28730,22 |
| Panel B: Portfolios as of July 2003 | | | | | |
| 1 | 13197 | 2535,42 | 250,20 | 603,20 | 1591,80 |
| 2 | 8927 | 4695,01 | 774,53 | 1646,20 | 3596,42 |
| 3 | 6815 | 7207,04 | 1444,63 | 2815,98 | 5973,14 |
| 4 | 5320 | 8368,42 | 2167,75 | 4178,13 | 8376,70 |
| 5 | 4097 | 12819,20 | 3099,98 | 5697,86 | 11416,66 |
| 6 to 9 | 10123 | 18788,08 | 5371,32 | 9649,33 | 18374,48 |
| More than 10 | 10307 | 50762,58 | 14457,45 | 26714,47 | 50355,61 |
| **All** | 58786 | 15903,96 | 1248,30 | 4241,70 | 13318,64 |
| Panel B: Portfolios as of July 2006 | | | | | |
| 1 | 10426 | 4295,99 | 390,40 | 1003,00 | 2740,20 |
| 2 | 6913 | 8637,35 | 1279,61 | 2815,50 | 6401,13 |
| 3 | 5240 | 12643,92 | 2443,90 | 5106,74 | 11053,84 |
| 4 | 3987 | 17353,15 | 3943,36 | 7437,04 | 16041,48 |
| 5 | 3158 | 22100,53 | 5327,29 | 9986,42 | 19985,22 |
| 6 to 9 | 7678 | 33021,51 | 9095,00 | 16809,07 | 33416,07 |
| More than 10 | 8096 | 88715,78 | 24546,38 | 46556,67 | 92137,80 |
| **All** | 45498 | 28166,44 | 2038,15 | 7324,47 | 23801,08 |

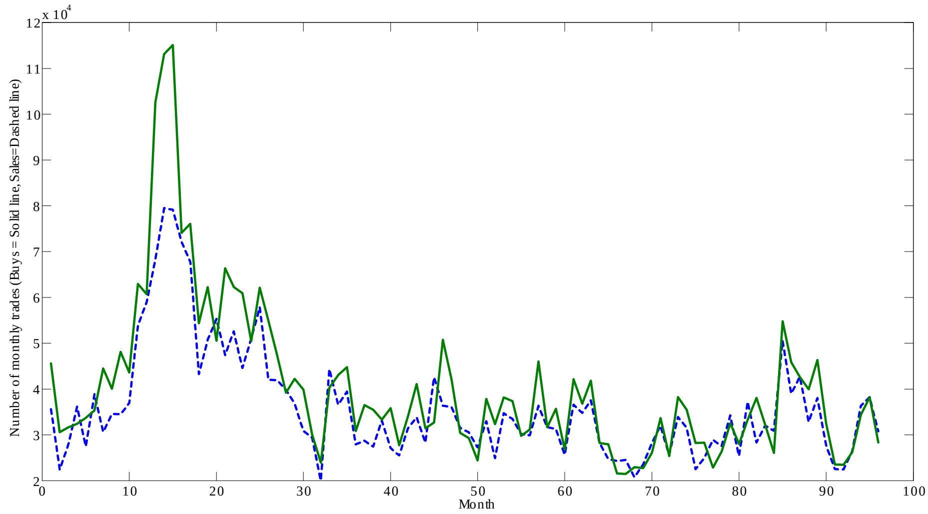


**Figure 4**
Time-series of the number of monthly trades. The solid (dashed) line represents the evolution of purchases (sales)

**Table 4**
Trades and holdings for stocks of the 6 main countries. FR = France, NL = The Netherlands, US = United States, GB = United Kingdom, DE = Germany, IT = Italy.

| 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**

This table provides the return statistics of investor portfolios for 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 ($\sigma$) are annualized. All moments, including skewnes ($S_k$) are estimated on one quarter of daily data

| | Subsequent returns sorted on $D_1$ | | | | | Subsequent returns sorted on $D_2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| July 00 | $N$ | $D_1$ | $R$ | $\sigma$ | $S_k$ | $N$ | $D_2$ | $R$ | $\sigma$ | $S_k$ |
| H Div. | 4055 | 0.044 | -0.071 | 0.186 | 0.072 | 4560 | 0.074 | -0.052 | 0.174 | 0.096 |
| 2 | 4003 | 0.074 | -0.085 | 0.205 | 0.095 | 4560 | 0.125 | -0.081 | 0.202 | 0.126 |
| 3 | 4535 | 0.102 | -0.100 | 0.220 | 0.114 | 4560 | 0.170 | -0.093 | 0.221 | 0.134 |
| 4 | 4622 | 0.135 | -0.109 | 0.234 | 0.120 | 4560 | 0.218 | -0.111 | 0.237 | 0.140 |
| 5 | 3194 | 0.167 | -0.118 | 0.247 | 0.139 | 4560 | 0.273 | -0.117 | 0.254 | 0.162 |
| 6 | 3640 | 0.200 | -0.111 | 0.255 | 0.153 | 4560 | 0.342 | -0.121 | 0.273 | 0.166 |
| 7 | 4643 | 0.250 | -0.123 | 0.274 | 0.182 | 4566 | 0.431 | -0.165 | 0.297 | 0.163 |
| 8 | 5730 | 0.333 | -0.141 | 0.291 | 0.176 | 4554 | 0.528 | -0.149 | 0.316 | 0.189 |
| 9 | 6618 | 0.500 | -0.170 | 0.329 | 0.214 | 4560 | 0.739 | -0.259 | 0.381 | 0.141 |
| L Div. | 7509 | 1.000 | -0.218 | 0.401 | 0.255 | 7509 | 1.000 | -0.218 | 0.401 | 0.255 |
| July 03 | $N$ | $D_1$ | $R$ | $\sigma$ | $S_k$ | $N$ | $D_2$ | $R$ | $\sigma$ | $S_k$ |
| H Div. | 4124 | 0.046 | 0.436 | 0.187 | 0.317 | 4951 | 0.081 | 0.454 | 0.185 | 0.313 |
| 2 | 4621 | 0.080 | 0.428 | 0.202 | 0.310 | 4952 | 0.137 | 0.438 | 0.204 | 0.308 |
| 3 | 5631 | 0.114 | 0.419 | 0.214 | 0.316 | 4951 | 0.186 | 0.433 | 0.217 | 0.304 |
| 4 | 2659 | 0.143 | 0.419 | 0.224 | 0.324 | 4952 | 0.239 | 0.429 | 0.228 | 0.318 |
| 5 | 3250 | 0.167 | 0.408 | 0.229 | 0.324 | 4951 | 0.297 | 0.403 | 0.236 | 0.339 |
| 6 | 4032 | 0.200 | 0.417 | 0.238 | 0.343 | 4952 | 0.364 | 0.409 | 0.252 | 0.359 |
| 7 | 5164 | 0.250 | 0.402 | 0.248 | 0.354 | 4951 | 0.457 | 0.376 | 0.260 | 0.373 |
| 8 | 6612 | 0.333 | 0.407 | 0.263 | 0.382 | 4952 | 0.539 | 0.383 | 0.277 | 0.410 |
| 9 | 8470 | 0.500 | 0.380 | 0.288 | 0.429 | 4951 | 0.728 | 0.316 | 0.286 | 0.470 |
| L Div. | 11403 | 1.000 | 0.331 | 0.330 | 0.536 | 11403 | 1.000 | 0.331 | 0.330 | 0.536 |
| July 06 | $N$ | $D_1$ | $R$ | $\sigma$ | $S_k$ | $N$ | $D_2$ | $R$ | $\sigma$ | $S_k$ |
| H Div. | 3343 | 0.046 | 0.151 | 0.148 | -0.092 | 3776 | 0.074 | 0.150 | 0.147 | -0.100 |
| 2 | 3541 | 0.080 | 0.164 | 0.158 | -0.033 | 3775 | 0.129 | 0.163 | 0.158 | -0.052 |
| 3 | 2516 | 0.106 | 0.162 | 0.164 | -0.021 | 3776 | 0.179 | 0.160 | 0.167 | -0.033 |
| 4 | 3723 | 0.135 | 0.164 | 0.171 | -0.015 | 3775 | 0.234 | 0.156 | 0.175 | -0.002 |
| 5 | 2389 | 0.167 | 0.165 | 0.175 | 0.032 | 3776 | 0.297 | 0.164 | 0.184 | 0.018 |
| 6 | 3083 | 0.200 | 0.158 | 0.183 | 0.026 | 3775 | 0.368 | 0.156 | 0.193 | 0.039 |
| 7 | 3862 | 0.250 | 0.160 | 0.190 | 0.059 | 3776 | 0.468 | 0.149 | 0.202 | 0.070 |
| 8 | 5036 | 0.333 | 0.148 | 0.202 | 0.060 | 3775 | 0.553 | 0.138 | 0.213 | 0.100 |
| 9 | 6487 | 0.500 | 0.133 | 0.221 | 0.120 | 3776 | 0.765 | 0.157 | 0.227 | 0.205 |
| L Div. | 8815 | 1.000 | 0.084 | 0.264 | 0.178 | 8815 | 1.000 | 0.084 | 0.264 | 0.178 |

**Table 6**
This table provides the return statistics of investor portfolios for 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 ($\sigma$) are annualized. All moments, including skewnes ($S_k$) are estimated on one quarter of daily data

| | Past returns sorted on $D_1$ | | | | | Past returns sorted on $D_2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| July 00 | $N$ | $D_1$ | $R$ | $\sigma$ | $S_k$ | $N$ | $D_2$ | $R$ | $\sigma$ | $S_k$ |
| H Div. | 4033 | 0.044 | 0.057 | 0.263 | -0.086 | 4383 | 0.073 | 0.066 | 0.239 | -0.113 |
| 2 | 3976 | 0.074 | 0.067 | 0.280 | -0.046 | 4384 | 0.123 | 0.065 | 0.272 | -0.033 |
| 3 | 4486 | 0.102 | 0.058 | 0.299 | -0.006 | 4383 | 0.167 | 0.049 | 0.291 | 0.001 |
| 4 | 4516 | 0.135 | 0.050 | 0.315 | 0.030 | 4384 | 0.214 | 0.043 | 0.310 | 0.035 |
| 5 | 3102 | 0.167 | 0.033 | 0.328 | 0.051 | 4383 | 0.267 | 0.028 | 0.332 | 0.062 |
| 6 | 3553 | 0.200 | 0.036 | 0.341 | 0.077 | 4384 | 0.335 | 0.026 | 0.356 | 0.081 |
| 7 | 4412 | 0.250 | -0.006 | 0.364 | 0.088 | 4383 | 0.420 | -0.009 | 0.388 | 0.089 |
| 8 | 5394 | 0.333 | 0.002 | 0.384 | 0.113 | 4384 | 0.526 | -0.007 | 0.395 | 0.144 |
| 9 | 5979 | 0.500 | -0.034 | 0.423 | 0.154 | 4383 | 0.738 | -0.063 | 0.496 | 0.174 |
| L Div. | 6666 | 1.000 | 0.002 | 0.484 | 0.205 | 6666 | 1.000 | 0.002 | 0.484 | 0.205 |
| July 03 | $N$ | $D_1$ | $R$ | $\sigma$ | $S_k$ | $N$ | $D_2$ | $R$ | $\sigma$ | $S_k$ |
| H Div. | 4131 | 0.046 | 0.978 | 0.232 | -0.033 | 5006 | 0.081 | 0.957 | 0.230 | -0.040 |
| 2 | 4644 | 0.080 | 0.964 | 0.252 | 0.023 | 5007 | 0.138 | 0.966 | 0.256 | 0.021 |
| 3 | 5660 | 0.114 | 0.958 | 0.266 | 0.053 | 5006 | 0.187 | 0.967 | 0.270 | 0.058 |
| 4 | 2672 | 0.143 | 0.944 | 0.279 | 0.083 | 5007 | 0.240 | 0.956 | 0.286 | 0.108 |
| 5 | 3268 | 0.167 | 0.932 | 0.285 | 0.111 | 5006 | 0.299 | 0.946 | 0.297 | 0.134 |
| 6 | 4067 | 0.200 | 0.949 | 0.296 | 0.140 | 5007 | 0.366 | 0.956 | 0.319 | 0.178 |
| 7 | 5250 | 0.250 | 0.931 | 0.311 | 0.161 | 5006 | 0.459 | 0.940 | 0.332 | 0.228 |
| 8 | 6717 | 0.333 | 0.930 | 0.333 | 0.212 | 5007 | 0.540 | 0.910 | 0.357 | 0.239 |
| 9 | 8649 | 0.500 | 0.891 | 0.369 | 0.279 | 5006 | 0.729 | 0.856 | 0.370 | 0.294 |
| Low Div. | 11585 | 1.000 | 0.755 | 0.413 | 0.343 | 11585 | 1.000 | 0.755 | 0.413 | 0.343 |
| July 06 | $N$ | $D_1$ | $R$ | $\sigma$ | $S_k$ | $N$ | $D_2$ | $R$ | $\sigma$ | $S_k$ |
| H Div. | 3372 | 0.046 | -0.171 | 0.213 | -0.376 | 3816 | 0.074 | -0.175 | 0.214 | -0.383 |
| 2 | 3571 | 0.080 | -0.160 | 0.218 | -0.336 | 3816 | 0.129 | -0.171 | 0.221 | -0.340 |
| 3 | 2542 | 0.106 | -0.169 | 0.223 | -0.312 | 3816 | 0.179 | -0.180 | 0.224 | -0.307 |
| 4 | 3762 | 0.135 | -0.176 | 0.227 | -0.295 | 3816 | 0.234 | -0.178 | 0.230 | -0.276 |
| 5 | 2421 | 0.167 | -0.172 | 0.229 | -0.275 | 3817 | 0.297 | -0.179 | 0.236 | -0.258 |
| 6 | 3120 | 0.200 | -0.169 | 0.234 | -0.246 | 3816 | 0.369 | -0.179 | 0.242 | -0.215 |
| 7 | 3897 | 0.250 | -0.172 | 0.241 | -0.213 | 3816 | 0.470 | -0.181 | 0.249 | -0.184 |
| 8 | 5073 | 0.333 | -0.184 | 0.250 | -0.183 | 3816 | 0.555 | -0.182 | 0.258 | -0.157 |
| 9 | 6587 | 0.500 | -0.176 | 0.263 | -0.127 | 3816 | 0.767 | -0.106 | 0.268 | -0.080 |
| L Div. | 9100 | 1.000 | -0.084 | 0.295 | -0.019 | 9100 | 1.000 | -0.084 | 0.295 | -0.019 |

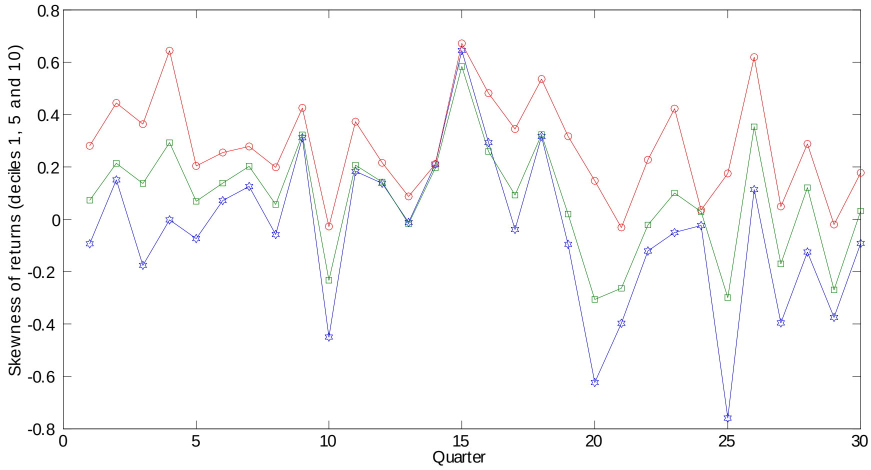


**Figure 5**
Evolution of median portfolio skewness over the period 1999-2006 for deciles 1 (high diversification identified by stars), 5 (squares) and 10 (low diversification, circles) for subsequent returns

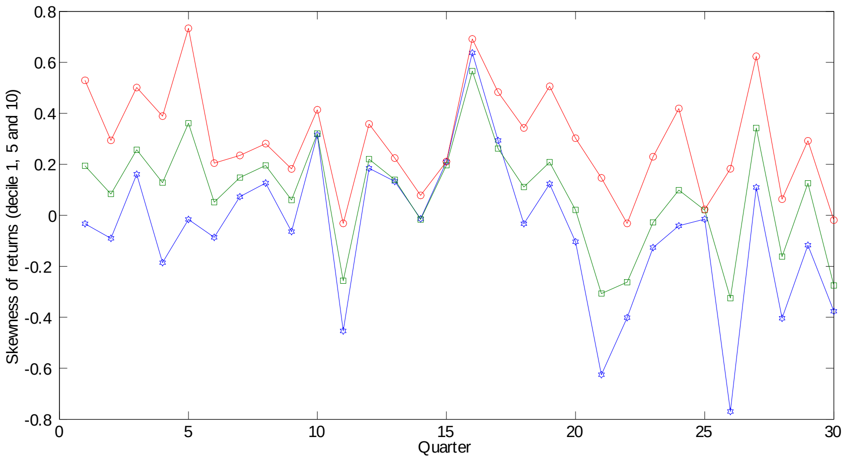


**Figure 6**
Evolution of median portfolio skewness over the period 1999-2006 for deciles 1 (high diversification identified by stars), 5 (squares) and 10 (low diversification, circles) for past returns

**Table 7**
This table provides the regression results of diversification measures ($D_1$ on the left of the table and $D_2$ on the right) on portfolio value and past skewness of returns. The regression equation is:

$$D = a_0 + a_P PortfolioValue + a_S Skewness + \varepsilon_i$$

For each diversification measure and each semester, the first column gives the number of investors and the adjusted $R^2$ of the regression, the three next columns provide the intercept and the coefficients of portfolio value and skewness. t-stats appear in the second line.

| | Panel A: Diversification $D_1$ | | | | Panel B: Diversification $D_2$ | | | |
|---|---|---|---|---|---|---|---|---|
| Semester | $N$ / $\overline{R}^2$ | $a_0$ | $a_P$ | $a_S$ | $N$ / $\overline{R}^2$ | $a_0$ | $a_P$ | $a_S$ |
| S1-99 | 30268 | 1.539 | -0.132 | 0.100 | 30268 | 1.405 | -0.108 | 0.143 |
| t-stat | 0.47 | 159.925 | -133.234 | 31.040 | 0.39 | 130.469 | -95.023 | 36.205 |
| S2-99 | 35311 | 1.590 | -0.135 | 0.019 | 35311 | 1.493 | -0.115 | 0.030 |
| t-stat | 0.46 | 178.739 | -147.644 | 9.472 | 0.35 | 153.196 | -111.017 | 13.722 |
| S1-00 | 39682 | 1.455 | -0.123 | 0.062 | 39682 | 1.352 | -0.105 | 0.097 |
| t-stat | 0.44 | 168.869 | -144.552 | 25.078 | 0.35 | 142.134 | -108.559 | 31.313 |
| S2-00 | 46688 | 1.495 | -0.128 | 0.010 | 46688 | 1.437 | -0.112 | 0.003 |
| t-stat | 0.44 | 207.035 | -171.185 | 4.002 | 0.34 | 186.328 | -134.047 | 1.135 |
| S1-01 | 52405 | 1.534 | -0.134 | 0.023 | 52405 | 1.450 | -0.115 | 0.030 |
| t-stat | 0.50 | 258.410 | -210.181 | 9.976 | 0.39 | 221.067 | -155.475 | 12.516 |
| S2-01 | 53151 | 1.579 | -0.139 | 0.023 | 53151 | 1.522 | -0.124 | 0.022 |
| t-stat | 0.50 | 262.876 | -210.541 | 16.079 | 0.42 | 246.179 | -175.073 | 14.675 |
| S1-02 | 52227 | 1.568 | -0.141 | 0.026 | 52227 | 1.491 | -0.121 | 0.027 |
| t-stat | 0.53 | 294.651 | -238.096 | 13.803 | 0.42 | 270.091 | -188.064 | 14.798 |
| S2-02 | 49484 | 1.550 | -0.142 | 0.033 | 49484 | 1.499 | -0.125 | 0.036 |
| t-stat | 0.49 | 249.361 | -210.460 | 11.060 | 0.42 | 239.827 | -179.472 | 11.925 |
| S1-03 | 48925 | 1.535 | -0.139 | 0.059 | 48925 | 1.490 | -0.124 | 0.064 |
| t-stat | 0.50 | 245.200 | -203.014 | 25.575 | 0.44 | 237.724 | -176.243 | 27.017 |
| S2-03 | 53970 | 1.560 | -0.138 | 0.049 | 53970 | 1.521 | -0.125 | 0.057 |
| t-stat | 0.50 | 271.441 | -222.920 | 24.858 | 0.45 | 266.398 | -196.930 | 28.452 |
| S1-04 | 51920 | 1.577 | -0.136 | 0.066 | 51920 | 1.538 | -0.123 | 0.077 |
| t-stat | 0.52 | 279.834 | -211.981 | 29.378 | 0.47 | 275.529 | -186.321 | 32.323 |
| S2-04 | 49537 | 1.600 | -0.140 | 0.038 | 49537 | 1.564 | -0.126 | 0.043 |
| t-stat | 0.50 | 275.334 | -219.170 | 21.894 | 0.45 | 271.335 | -193.705 | 24.269 |
| S1-05 | 47576 | 1.624 | -0.140 | 0.020 | 47576 | 1.585 | -0.126 | 0.031 |
| t-stat | 0.51 | 289.312 | -227.429 | 14.064 | 0.46 | 286.617 | -200.785 | 21.651 |
| S2-05 | 44950 | 1.609 | -0.137 | 0.037 | 44950 | 1.571 | -0.124 | 0.052 |
| t-stat | 0.51 | 267.277 | -213.454 | 23.874 | 0.46 | 265.819 | -189.996 | 32.589 |
| S1-06 | 42040 | 1.597 | -0.135 | 0.040 | 42040 | 1.568 | -0.122 | 0.050 |
| t-stat | 0.51 | 264.828 | -209.195 | 21.935 | 0.45 | 266.155 | -186.460 | 26.583 |
| S2-06 | 40158 | 1.580 | -0.132 | 0.044 | 40158 | 1.541 | -0.119 | 0.057 |
| t-stat | 0.49 | 242.984 | -196.249 | 22.699 | 0.44 | 240.348 | -173.153 | 27.806 |

**Table 8**
This table provides on panel A (B) the Spearman rank correlation between the SAID calculated with $D_1$ and the skewness of one year of daily past (subsequent) returns. $N$ is the number of stocks held at the end of the corresponding year. Rho is the Spearman rank correlation coefficient between $D_1$ 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)

| | Panel A: Past returns. $D_1$ | | | | | | Panel B: Subsequent returns. $D_1$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | $\rho$ | pval | Hdiv | Mdiv | Ldiv | N | $\rho$ | pval | Hdiv | Mdiv | Ldiv |
| 1999 | 1042 | 0.02 | 0.60 | 0.36 | 0.58 | 0.51 | 1141 | 0.06 | 0.04 | 0.34 | 0.61 | 0.46 |
| 2000 | 1232 | 0.03 | 0.24 | 0.34 | 0.49 | 0.46 | 1315 | 0.11 | 0.00 | 0.18 | 0.28 | 0.45 |
| 2001 | 1318 | 0.08 | 0.00 | 0.21 | 0.22 | 0.43 | 1315 | 0.08 | 0.00 | 0.17 | 0.20 | 0.39 |
| 2002 | 1296 | 0.06 | 0.02 | 0.14 | 0.30 | 0.32 | 1221 | 0.04 | 0.20 | 0.52 | 0.39 | 0.40 |
| 2003 | 1239 | 0.02 | 0.49 | 0.53 | 0.41 | 0.40 | 1189 | -0.04 | 0.13 | 0.34 | 0.35 | 0.17 |
| 2004 | 1196 | -0.03 | 0.29 | 0.41 | 0.44 | 0.29 | 1142 | -0.01 | 0.69 | 0.58 | 0.41 | 0.53 |
| 2005 | 1182 | 0.02 | 0.51 | 0.52 | 0.43 | 0.31 | 1165 | 0.01 | 0.86 | 0.52 | 0.50 | 0.39 |
| 2006 | 1143 | 0.01 | 0.63 | 0.34 | 0.52 | 0.40 | 1169 | 0.01 | 0.78 | 0.43 | 0.37 | 0.26 |

**Table 9**
This table provides on panel A (B) the Spearman rank correlation between the SAID calculated with $D_2$ and the skewness of one year of daily past (subsequent) returns. $N$ is the number of stocks held at the end of the corresponding year. Rho is the Spearman rank correlation coefficient between $D_2$ 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)

| | Panel A: Past returns. $D_2$ | | | | | | Panel B: Subsequent returns. $D_2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | $\rho$ | pval | Hdiv | Mdiv | Ldiv | N | $\rho$ | pval | Hdiv | Mdiv | Ldiv |
| 1999 | 1042 | 0.04 | 0.23 | 0.37 | 0.70 | 0.66 | 1141 | 0.08 | 0.01 | 0.32 | 0.45 | 0.47 |
| 2000 | 1232 | 0.06 | 0.04 | 0.34 | 0.49 | 0.45 | 1315 | 0.13 | 0.00 | 0.27 | 0.45 | 0.41 |
| 2001 | 1318 | 0.12 | 0.00 | 0.16 | 0.37 | 0.43 | 1315 | 0.11 | 0.00 | 0.16 | 0.25 | 0.35 |
| 2002 | 1296 | 0.08 | 0.00 | 0.14 | 0.24 | 0.33 | 1221 | 0.06 | 0.05 | 0.47 | 0.35 | 0.42 |
| 2003 | 1239 | 0.03 | 0.35 | 0.45 | 0.46 | 0.40 | 1189 | -0.05 | 0.09 | 0.27 | 0.40 | 0.07 |
| 2004 | 1196 | -0.05 | 0.07 | 0.44 | 0.44 | 0.15 | 1142 | 0.01 | 0.81 | 0.55 | 0.40 | 0.48 |
| 2005 | 1182 | 0.01 | 0.64 | 0.50 | 0.49 | 0.31 | 1165 | 0.02 | 0.51 | 0.36 | 0.53 | 0.31 |
| 2006 | 1143 | 0.04 | 0.21 | 0.32 | 0.53 | 0.40 | 1169 | 0.03 | 0.32 | 0.42 | 0.28 | 0.23 |

**Table 10**
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 11**
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 the1% 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 the10% level | 96 | 91 | 88 | 87 | 86 |
| Panel B: 102 stocks (average) randomly chosen | | | | | |
| Percentage of significant rank correlations at the1% 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 the10% level | 99 | 95 | 95 | 93 | 93 |
| Panel B: 162 stocks (average) randomly chosen | | | | | |
| Percentage of significant rank correlations at the1% 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 the10% level | 99 | 98 | 99 | 97 | 97 |