Rev Quant Finan Acc (2016) 47:129–157
DOI 10.1007/s11156-015-0497-1

ORIGINAL RESEARCH

Diversification, gambling and market forces

Marie-Hélène Broihanne · Maxime Merli · Patrick Roger

Published online: 29 January 2015

Abstract Though simple and appealing, mean-variance portfolio choice theory does not describe actual diversification choices by investors, especially their propensity to gamble and the solvency constraints they face. Using 8 million trades realized by 90,000 individual investors, we show that diversification choices are in fact strongly driven by the skewness of returns, especially in bull markets, but also by the amount to be invested in risky assets. Increasing this amount by 10 % leads to increase by 3.8 % the number of stocks in investors' portfolios, controlling for portfolio skewness. An important contribution of this paper is to show that the strength of the relationship between diversification and the skewness of returns is shaped by market forces. A strong negative relationship exists in bull markets but disappears in bear markets, a result not found in the literature. Our results survive several robustness checks, including controlling for individual heterogeneity and time-variability of stock price co-movements.

Keywords Individual investors · Return skewness · Diversification · Gambling

JEL Classification G02 · G11

1 Introduction

In an economy governed by the standard theory of portfolio choice, investors hold diversified portfolios. They are looking for a profile of returns characterized by a high first

M.-H. Broihanne · M. Merli · P. Roger (✉)
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

M.-H. Broihanne
e-mail: mhb@unistra.fr

M. Merli
e-mail: merli@unistra.fr

moment and a low second moment. However, retail investors actually hold underdiversified portfolios (generally with less than five stocks). Most of the time, these investors hold positively skewed portfolios (Mitton and Vorkink 2007).

Several studies justify the underdiversification of retail investors portfolios. A first reason of underdiversification is the investors' desire for positive skewness (Barberis and Huang 2008; Brunnermeier et al. 2007; Brunnermeier and Parker 2005). A second reason is that retail investors bear solvency constraints and transaction costs (Liu 2014). In such a constrained environment, Liu shows that diversification choices are driven by the amount to be invested in risky assets. Less wealthy investors hold a portfolio with one or two stocks, and more generally, the optimal number of stocks for them is an increasing function of the amount invested in risky assets. It turns out that, in this approach, the positive skewness of returns is only a by-product of underdiversification.

The purpose of this paper is to reconcile the two above approaches. We first show, by theoretical results and simulations that, in an Arrow–Debreu economy, the skewness of a portfolio return decreases with the number of stocks in the portfolio.[1] We also run a Chamberlain and Rothschild (1983) multi-factor model, and show that the strength of the link between skewness of portfolio returns and diversification comes essentially from the aggregate share of variance on common factors (called market forces in the title of the paper).

Second, we analyze the trading records and portfolios of 87,373 individual investors over the period 1999–2006. We show that the main variables explaining diversification choices are the amount invested in risky assets (called "portfolio value" hereafter) and the skewness of returns. We stress the strong impact of portfolio value on diversification choices of retail investors. Therefore, our empirical results are in line with the predictions of Liu's model (2014). When controlling for portfolio value, we show that the importance of skewness differs considerably between bull periods and bear periods. In bull periods, the relationship between skewness of returns and diversification is similar to (a) the relationship found by Mitton and Vorkink (2007) and (b) our theoretical result on Arrow–Debreu markets. However, in bear periods (especially 2001–2002), the relationship between skewness of returns and diversification is non-significant. In these periods, diversification choices are almost entirely explained by the portfolio value. Investors' desire for skewness in returns is no longer there. In particular, our results show that in bear markets the relationship between diversification and skewness is reversed.[2] Therefore, our empirical study contributes to the literature by showing how market forces help to understand precisely the (dual) causal relationships between diversification and skewness.

This paper is organized as follows. Section 2 reviews the literature on portfolios' (under-) diversification and skewness-seeking by investors. In Sect. 3, we present the theoretical and simulation results in Arrow–Debreu markets. Section 4 describes our database and Sect. 5 details our empirical results. A short conclusion gives recommendations for future research.

[1] Harvey and Siddique (1999, 2000) and Chen et al. (2001) show that the average skewness of single stocks is positive in most periods and the market skewness is negative most of the time. More recently, Albuquerque (2012) got the same results except during the second half of 1987 (due to the Black Monday). The skewness of the equally-weighted market portfolio is negative 77 % of the time.

[2] At the same time, diversification does not reduce by much the portfolio variance because systematic risk is the most important component of total risk in such periods.

2 Related literature

The underdiversification of portfolios held by retail investors was first highlighted by Lease et al. (1974) and Blume and Friend (1975), followed by Kelly (1995). More recently, a number of empirical studies (Kumar 2007; Goetzmann and Kumar 2007; Mitton and Vorkink 2007; Odean 1999) have also shown in large samples that, individual investors hold largely underdiversified portfolios, containing less than five stocks on average.[3]

Two psychological traits justify the investors' desire for positive skewness in portfolio returns. First, the propensity to gamble leads to positively skewed portfolios in the hope of obtaining 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 (Bali et al. 2011; Kumar 2009). Second, preferring high skewness may simply reveal a prudent behavior in the sense discussed by Kimball (1990). In this case, investors are said to be downside risk-averse (Eeckhoudt and Schlesinger 2006; Menezes et al. 1980) 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 thus compatible with positive skewness-seeking (Deck and Schlesinger 2010; Ebert and Wiesen 2011; Tarazona-Gomez 2004). If underdiversification is a way to capture high skewness, investors who choose their optimal portfolios by considering the three first moments of the distribution of returns, would build underdiversified portfolios. In fact, one of the first attempts to introduce the third moment of the distribution of returns into a portfolio choice model was proposed by Kraus and Litzenberger (1976), followed by Harvey and Siddique (2000), who provided empirical support for this model. Mitton and Vorkink (2007) also based their theoretical analysis on a three-moment model and showed that heterogeneity in the preference for skewness induces underdiversification at equilibrium. More recently, Conrad et al. (2013) also show that the idiosyncratic skewness of returns is priced. Ortobelli et al. (2005) and Stoyanov et al. (2011) and study portfolio choice with fat-tailed distributions of returns [see also Francis and Kim (2013)] and Kim et al. (2014) proposed a robust mean-variance approach that can control portfolio skewness.

In the framework of non-expected utility models, desirability of positive skewness appears 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 positively skewed because they are 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). Investors' utility functions are concave for gains and convex for losses. Moreover, investors distort probabilities to overweight extreme outcomes. When a positively skewed asset is traded, 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, leading investors to avoid large losses and to search for large gains, even if the corresponding objective probability is very low. Again, such preferences lead investors to build positively skewed portfolios.

[3] Calvet et al. (2007) obtained the same results for Sweden, except that Swedish investors seem to have slightly more diversified portfolios than U.S. investors. Concerning the performance of individual investors, see for example Barber and Odean (2000), Shu et al. (2004), Entrop et al. (2014).

One potential drawback of CPT is the fact that the transformation of probabilities is “exogenous”; it depends only 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 upon the first draw. Upon the second draw, the potential outcomes are almost identical, except that the jackpot increases 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.

To overcome this drawback, Brunnermeier and Parker (2005) went one step further by considering the distorted probability measure as a decision variable; they call this probability measure *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. Moreover, the optimal distortion of the probability measure not only depends on the ranking of outcomes, but on the value of outcomes. In the lottery example of the above paragraph, the probability distortion is different with, and without, a rollover of the jackpot.

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 states.

It is thus 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 whereas 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), which 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”, 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; Forrest et al. 2002; Garrett and Sobel 1999; Walker and Young 2001). On stock markets, this effect has been recently illustrated by Kumar (2009) and Bali et al. (2011). Kumar (2009) shows that for some categories of individual investors, there is a strong link between portfolio choice and behavior on gambling markets such as 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 et al. (2011) 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 demonstrate the persistence of this ranking over time by calculating transition probability matrices from one month to the next. Thirty-five percent of stocks in the highest decile 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 into the decision process. In fact, diversification reduces (undesirable) idiosyncratic volatility but also reduces (desirable) positive skewness.

3 Model and simulations: the case of Arrow–Debreu markets

In this section, we start by developing analytical formulae (reported in the Appendix) for the first three moments of equally weighted portfolios of Arrow–Debreu (AD) securities and link these formulas to typical measures of diversification. In the next subsection, simulations allow us to show that these links hold true in more general frameworks.

3.1 Portfolios of Arrow–Debreu securities

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

Table 1 shows the evolution of variance and the third central moment as a function of the number of AD securities in portfolios. We assume a finite state-space with 20 equally-likely states of nature (i.e. $n = 20$). The third moment becomes slightly negative when the number of AD securities is greater than 10, but the variance decreases more rapidly. This finding implies that standardized skewness, defined as s_k^3/σ_k^3, becomes largely negative when the portfolio is sufficiently diversified. 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 the skewness of a single stock with a time series of returns, one is

Table 1 Second and third moments as a function of the number of Arrow–Debreu securities in the portfolio

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			

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 portfolios containing 11 to 19 Arrow–Debreu securities likely to observe 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 arrive for all stocks in the portfolio. On the contrary, it is well documented that correlations of stock returns increase in hard times.[4] Consequently, low returns are more likely to be observed simultaneously, leading to negative skewness for portfolio returns (Albuquerque 2012). A "similar" phenomenon is observed for portfolios of Arrow–Debreu securities. The payoff of a single AD is typically positively skewed, paying one in one state and 0 otherwise. An equally weighted portfolio of a large number (say k) of AD securities pays $1/k$ on k states and 0 on the remaining $n-k$ states. As k approaches n, the portfolio payoff becomes negatively skewed because losses are less likely but larger (for a given initial investment).

To verify whether this result holds 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 $x_1, \ldots, x_m$ between 0 and 100 and define the weights as $w_k = x_k / \sum_{j=1}^{m} x_j$. We consider only positive weights because individual investors almost never use short selling.

For each m, we calculate the average skewness and the 99 % confidence bounds (5th lowest and 5th highest skewness in a sample of 1,000 portfolios).

Figure 1 shows the evolution of the average skewness as well as the confidence bounds (dashed lines). There is no uncertainty for single-stock portfolios because of the analytical solution presented in Proposition 1 (Appendix). The average skewness of simulated portfolios is always positive whatever the number of stocks in portfolios, reaching close to 0 only for the maximum number of stocks. At first glance, this result could seem surprising because, in the equally weighted case of the preceding section, the skewness became negative when $k > n/2$. This result is attributed to the fact that portfolio weights are simulated according to a uniform multidimensional distribution. Consequently, the proportion of portfolios close to the equally weighted case is very low, and these are the portfolios with the lowest skewness. The bold curve in Fig. 1 is comparable to the results obtained by Mitton and Vorkink (2007) (Table 3, p. 1271); that is, the curve shows a similar evolution of skewness with respect to diversification. Consequently, it appears that underdiversification is a good way to capture high skewness. It is even unnecessary to pick highly skewed stocks to obtain a highly skewed portfolio. There is a "mechanical" relationship between diversification and skewness of returns.

3.2 The general case

In a finite state-space, a stock is a portfolio of Arrow–Debreu securities. To generalize the abovementioned results, we consider now 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. Denote this set of securities $E_k = \{e_1, \ldots, e_{n_k}\}$. We then define the payoffs of stock k by drawing n_k random weights $w_j, j = 1, \ldots, n_k$ (summing to one), as in the preceding section. The payoffs of stock k are equal to $\sum_{j=1}^{n_k} w_j e_j$ where e_j denotes the j-th AD

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

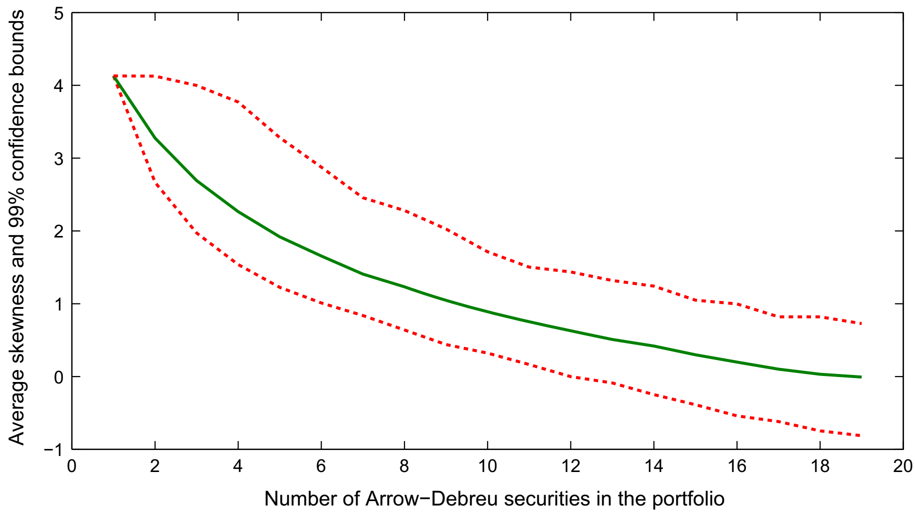

Fig. 1 Skewness versus diversification in an Arrow–Debreu world. The *solid line* gives the evolution of the average skewness of portfolios as a function of the number of Arrow–Debreu securities in portfolios. Each point is the average over 1,000 simulated markets. The *dashed lines* represent the corresponding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

security of E_k. This process is iterated n times and leads to a (n, n) matrix of payoffs of n single stocks, which corresponds to one simulated market.

Viewing the problem from a geometrical point of view is useful to understand what is the link between diversification and skewness. In the AD market, adding one AD security means adding one dimension to the space spanned by securities because AD securities are orthogonal vectors. However, typical stocks, i.e., portfolios of AD securities, may be correlated, and the number of dimensions spanned by stocks may be lower than the number of states. The probabilistic view of the same problem holds that when market (idiosyncratic) risk represents a large (small) share of total variance, stocks do not span a large space and diversification will not greatly reduce the skewness of portfolio returns (compared to the skewness of single stocks).

Simulation is an interesting way to check if our conjecture is true. Remember that empirical studies show that skewness decreases with diversification, but 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 0.2 in January 1991, 0.36 in January 1993 and, finally, 0.54 in January 1996.

To introduce our methodology, consider the one-factor CAPM-like model in which returns are written (with standard notations) as follows:

$$r_k = r_f + \beta_k (r_M - r_f) + \varepsilon_k \tag{1}$$

The variance of r_k, denoted σ_k^2, is equal to $\beta_k^2 \sigma_M^2 + \sigma(\varepsilon_k)^2$ where σ_M^2 is the variance of the market portfolio, and $\sigma(\varepsilon_k)^2$ is the idiosyncratic variance of stock k. An estimate of the ratio of market variance over total variance, calculated as $\sigma_M^2 \sum_{k=1}^n \beta_k^2 / \sum_{k=1}^n \sigma_k^2$, is given by the first eigenvalue of the covariance matrix of returns. It is calculated as a percentage of the trace of the covariance matrix which is equal to $\sum_{k=1}^{n} \sigma_k^2$. This estimation is based on the approximate factor structures proposed by Chamberlain and Rothschild (1983).

More generally, in a 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 thus that the average skewness decreases more rapidly 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. However, estimating the number of common factors by a principal component analysis is difficult because there is a bias toward the identification of a single-factor model in finite samples (Harding 2008). 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.

One thousand simulations are performed, and Table 2 summarizes the findings. Panel A (B, C) presents the main results for the case in which 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 $m = 1, \ldots, 5$. The second line provides the Spearman rank correlation between this percentage of variance and the decrease in skewness across simulations. These correlations are valued over 1,000 simulations; they are significant at all of the standard levels. In Panel A, the first factor aggregates (on average) 24.06 % of 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, which 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.

Table 2 Correlation between skewness decrease due to diversification and cumulated variance of common factors

Number of factors (k)	1	2	3	4	5
Panel A: 1,000 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: 1,000 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 C: 1,000 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

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 $k = 1, \ldots, 5$. 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

These 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, we mentioned before that bearish markets tend to increase the average correlation between stock returns (Ang and Bekaert 2002, 2004; Campbell et al. 2002; Longin and Solnik 2001), meaning that the percentage of variance linked to common factors is higher. We thus observe a lower decrease in skewness due to diversification. In the next section, we illustrate this point for our large sample of individual investors. In fact, it is worth mentioning that a low percentage of simulations leads to an increase in skewness (6.6, 2.4 and 2.7 %, respectively, in panels A, B and C) instead of a decrease when the number of stocks in the portfolio increases. Thus, the equally weighted portfolio is more positively skewed than the average stock in such infrequent cases. These cases correspond to high levels of market variance because 30.1 % (20.03, 16.8 %) of variance lies on the first factor, compared to an average of 24.06 % (14.27, 11.79 %) in Table 2.

4 Data, descriptive statistics and first results

4.1 Investors and stock data

Data on individual investors are derived from a large French brokerage house. We obtained transaction data for all active accounts over the period 1999–2006, amounting to a total of 9 million trades among 92,603 investors. Two different files are used in the present paper. First, the trades file provides the following information for each trade: the ISIN code of the asset, the buy-sell indicator, the date, the quantity and the amount in Euros. Second, the investors file compiles some characteristics of investors: date of birth, gender, date of entry to 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; others close their account before the end of the period. It would make no sense to analyze portfolios every day (due to the low turnover), so we chose to “take a photograph” of portfolios at the end of each month. As a result, some investors may hold no position in a given month, even if they held a portfolio before and resume trading 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 for at least two successive months throughout the period), but their number varies over time. Thus, 8,258,809 trades remain in our final database. On average, the number of investors in a month is 51,340, with a minimum of 34,230 and a maximum of 60,001.

Figure 2 shows three time series. The upper dotted curve represents the average number of stocks held by investors ($\times 10^4$). It varies from 5.5 to 6.8, and the median is 3 or 4 over 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 most striking feature of this curve is the sharp increase in the average number of stocks just before the dotcom bubble burst, which occurred during the first months of 2000. Then, a decrease to the former level of diversification is observed. Over 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 Goetzmann and

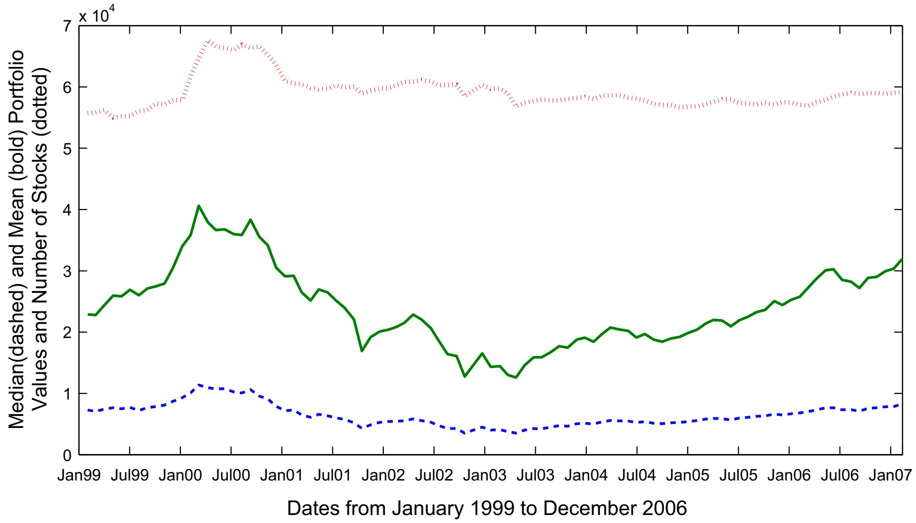

Fig. 2 Average number of stocks and portfolio values. The *three curves* represent respectively the time-series of the average number of stocks held by investors, and the mean and median portfolio value. The period under consideration starts in January 1999 (month 1) and ends in December 2006 (month 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

Kumar (2007) for a sample of U.S. investors. As mentioned previously, these authors observed an increase in diversification over the period 1991–1996 because the market was bullish over nearly the entire period.

The middle bold curve and the bottom dashed curves present, 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 overall market. 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 study period (December 2006). Consequently, the evolution of average portfolio values in our sample is not different from the evolution of the stock market.

Regarding portfolio value, there is a large discrepancy between the mean and median portfolio values, a result that is in line with other studies on individual investors. In fact, a few investors are very wealthy and invest a lot of money in risky assets, compared to the average investor; these wealthy investors shift the average portfolio value significantly upward. It is worth mentioning that 0.2 % of investors hold a stock portfolio worth more than one million euros.

Table 3 provides more detailed statistics at three points in time, July 2000, July 2003 and July 2006.[5] 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 presents the number of investors in

[5] We use the same presentation as that of Table 2 of Mitton and Vorkink (2007). The complete statistics for all months of the period are available upon request.

Table 3 Descriptive statistics on portfolio values at three points in time

Portfolio Value (in €)					
Number of stocks	Number of observations	Mean	1st quartile	Median	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 C: 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

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

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 (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. The result supports the preceding remarks made about Fig. 2. In most cases, the mean is close to the third quartile.

The market activity of investors in our sample is also highly variable over time. Figure 3 shows the time-series of monthly trades. The bold (dashed) line represents buy(sell) trades. The large variations are essentially observed in the three first years, with a dramatic

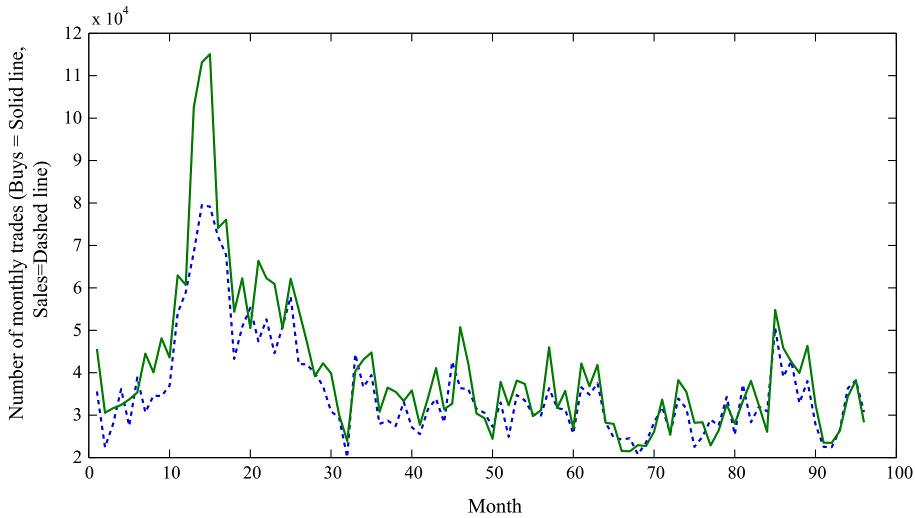

Fig. 3 Time-series of monthly trades. The *solid* (*dashed*) *line* represents the evolution of purchases (sales)

increase 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 then occurred until September 2001.

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

Price data are derived from two sources, Eurofidai for stocks traded on Euronext[6] and Bloomberg for the other stocks. The Eurofidai database provides price and return data for stocks traded in Europe. It is built in the spirit of the US database of the Center for Research on Security Prices (CRSP). We used daily prices to estimate the moments of the distribution of returns on stocks and investors portfolios. Our sample, contains 2,491 stocks, and 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 principally 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 account for more than 90 % of the trading volume, as shown in panel A of Table 4. The table illustrates the well-known home bias puzzle.[7] Thus, most comparisons in this paper are related to the French market. Moreover, the trading volume on U.S. stocks is very low. Only 54,881 trades on U.S. stocks were executed, compared for example to the 366,138 trades on 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 approximately 3 % held only foreign stocks), 21.05 % were holding Dutch stocks but only 3.97 % were holding U.S.

[6] 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. (2012) to study the contagion of sentiment across countries, including France and the U.S.

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

Table 4 Trades and holdings for stocks of the 6 main countries

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

	Number of investors	FR	NL	US	GB	DE	IT
Panel B: Percentage of investors holding stocks of the 6 main countries							
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

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

stocks, despite the large number of U.S. stocks in the database (that is, stocks traded at least once over the period).

4.2 Three measures of diversification

Despite our theoretical results on Arrow–Debreu markets presented in Sect. 3, it is unclear whether underdiversified portfolios should bear more skewness in real markets. Using three 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.[8] We use past returns to estimate skewness because, from a behavioral point of view, investors base their choices on the observation of past skewness.

The three diversification measures, denoted D_1, D_2 and D_3 are defined as follows.

D_1 is the inverse of the number of different stocks in the portfolio.

$$D_1^j = \frac{1}{n_j} \tag{2}$$

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

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

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

$$D_2^j = \sum_{i=1}^{n} w_{ij}^2 \tag{3}$$

where w_{ij} is the weight of security i in investor j's portfolio. D_2 also takes higher values at lower levels of diversification.

Finally, the third measure, denoted D_3 aims at taking into account correlations between stock returns. It is the ratio of the portfolio variance over the average variance of individual stocks composing the portfolio.

$$D_3^j = \frac{n_j V_j}{\sum_{i=1}^{n_j} \sigma_i^2} \tag{4}$$

where σ_i^2 is the variance of the return of asset i and V_j is the variance of the portfolio held by investor j. D_3 is called "normalized variance" in Goetzmann and Kumar (2007).

The three measures are positively correlated because all of these indices take high values for underdiversified portfolios and low values for largely diversified portfolios.

4.3 Diversification and portfolio skewness

In this section, we perform an investor-level analysis.We analyze the link between underdiversification and portfolio skewness.[9] Though most of the time the link obtained through simulations also appears in real data, periods of sharp market drops do not reveal the same relationship. To study the link between diversification and skewness, we sort all investor portfolios into deciles according to measures of diversification D_1, D_2 and D_3 at the beginning of each quarter. We calculate the average value of $\widehat{S_k}^3$ within each decile, as well as average returns and standard deviations (which are annualized in Table 5). We cannot present the results for all quarters here, so we consider two examples; July 2002 illustrates a bearish period and July 2005 illustrates a bullish period.

Table 5 provides the estimates of moments and the values of diversification measures. Concerning D_1, the number of investors is different from one decile to another because D_1 is a discrete variable (the inverse of the number of stocks). Therefore, it is meaningless to arbitrarily allocate investors to different deciles when they have the same value of D_1. Concerning the Herfindahl index D_2, and the normalized variance D_3, this problem concerns only single-stock holders. In addition, D_3 can be greater than 1.

Whatever the diversification index at hand, there is a stable relationship between D_i, $i = 1, 2, 3$ and the second moment of the distribution of returns. The relationship between D_i and skewness is almost always increasing, except at three points in time corresponding to the period July 2001 to December 2002. Thus, a higher D indicates lower diversification,

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

$$\widehat{S_k}^3 = \frac{\frac{1}{n}\sum_{t=1}^{n}(r_t - \bar{r})^3}{\widehat{\sigma}^3} \tag{5}$$

where $\bar{r}$ is the average daily return and $\widehat{\sigma}^3$ the cube of the estimated standard deviation of daily returns. One advantage of Eq. (5) is that it is standardized by variance (or standard deviation).The equation offers a way to take into account the mechanical positive link between variance and skewness illustrated in Sect. 3.

Table 5 Portfolio returns and diversification during bullish and bearish periods

	July 2002: Past returns					July 2005: Past returns				
	D_1									
	N	D_1	R	σ	S_k	N	D_1	R	σ	S_k
H Div.	4,213	0.046	−0.579	0.234	0.218	3,395	0.046	0.273	0.118	−0.743
2	4,670	0.08	−0.598	0.254	0.211	3,856	0.08	0.274	0.128	−0.562
3	3,261	0.106	−0.606	0.265	0.19	4,599	0.114	0.264	0.136	−0.456
4	4,657	0.135	−0.62	0.276	0.181	2,331	0.143	0.257	0.143	−0.404
5	3,170	0.167	−0.626	0.291	0.197	2,765	0.167	0.253	0.148	−0.349
6	3,901	0.200	−0.629	0.299	0.174	3,392	0.200	0.254	0.151	−0.305
7	4,853	0.250	−0.641	0.314	0.177	4,341	0.250	0.244	0.159	−0.254
8	6,198	0.333	−0.666	0.339	0.154	5,679	0.333	0.251	0.168	−0.152
9	8,097	0.500	−0.675	0.376	0.158	7,545	0.500	0.242	0.183	−0.056
L Div.	10,685	1.000	−0.670	0.429	0.121	10,901	1.000	0.281	0.205	0.092
	D_2									
	N	D_2	R	σ	S_k	N	D_2	R	σ	S_k
H Div.	4,787	0.081	−0.56	0.23	0.183	4,215	0.077	0.27	0.119	−0.736
2	4,787	0.138	−0.603	0.255	0.182	4,215	0.134	0.264	0.13	−0.551
3	4,787	0.188	−0.616	0.271	0.178	4,214	0.184	0.257	0.139	−0.453
4	4,787	0.243	−0.639	0.29	0.178	4,215	0.238	0.252	0.145	−0.368
5	4,787	0.305	−0.645	0.303	0.188	4,215	0.301	0.245	0.153	−0.305
6	4,787	0.374	−0.662	0.322	0.176	4,215	0.370	0.240	0.160	−0.214
7	4,787	0.47	−0.659	0.331	0.167	4,214	0.469	0.247	0.17	−0.126
8	4,787	0.552	−0.681	0.366	0.179	4,215	0.551	0.245	0.176	−0.071
9	4,787	0.747	−0.603	0.337	0.212	4,215	0.755	0.306	0.181	0.065
Low Div.	10,622	1.000	−0.67	0.432	0.121	10,871	1.000	0.281	0.205	0.092
	D_3									
	N	D_3	R	σ	S_k	N	D_3	R	σ	S_k
H Div.	4,852	0.219	−0.509	0.207	0.052	4,265	0.174	0.272	0.131	−0.535
2	4,852	0.309	−0.583	0.246	0.151	4,265	0.265	0.273	0.135	−0.486
3	4,852	0.361	−0.618	0.271	0.182	4,266	0.332	0.269	0.136	−0.470
4	4,852	0.409	−0.637	0.29	0.186	4,265	0.392	0.261	0.141	−0.412
5	4,852	0.458	−0.644	0.309	0.194	4,265	0.453	0.246	0.146	−0.377
6	4,852	0.515	−0.663	0.334	0.212	4,265	0.519	0.25	0.152	−0.292
7	4,852	0.583	−0.67	0.357	0.226	4,266	0.594	0.237	0.161	−0.241
8	4853	0.673	−0.697	0.38	0.23	4,265	0.685	0.233	0.169	−0.200
9	4,851	0.839	−0.647	0.399	0.282	4,265	0.847	0.292	0.182	−0.036

Table 5 continued

	D_3									
	N	D_3	R	σ	S_k	N	D_3	R	σ	S_k
L Div.	10,037	1.000	−0.726	0.454	0.108	10,417	1,000	0.276	0.202	0.091

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 at the top of the table, D_2 in the middle and D_3 at the bottom). The left (right) part of the table shows returns as of July 02 (July 05). 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

higher variance and higher skewness. However, the relation between diversification and skewness is broken during crisis periods. In normal periods, we obtain results in line with our theoretical results of Sect. 3, and in line with the empirical findings of Mitton and Vorkink (2007). The difference between the behavior of stock prices in crisis periods and our theoretical results is attributed to the peculiarities of Arrow–Debreu securities: they are positively skewed, as are single stocks in normal periods. However, during crisis periods, a number of stocks can become negatively skewed. In this case, diversifying does not decrease the skewness of returns.

When looking at all quarters, skewness almost always decreases when diversification increases. However, over time, the mean level of skewness evolves according to market movements, and the same observation can be made regarding the variation of skewness between the first and the last deciles. Figure 4 presents the dynamics of median skewness of returns on deciles 1 (high diversification), 5 (medium diversification) and 10 (low diversification).

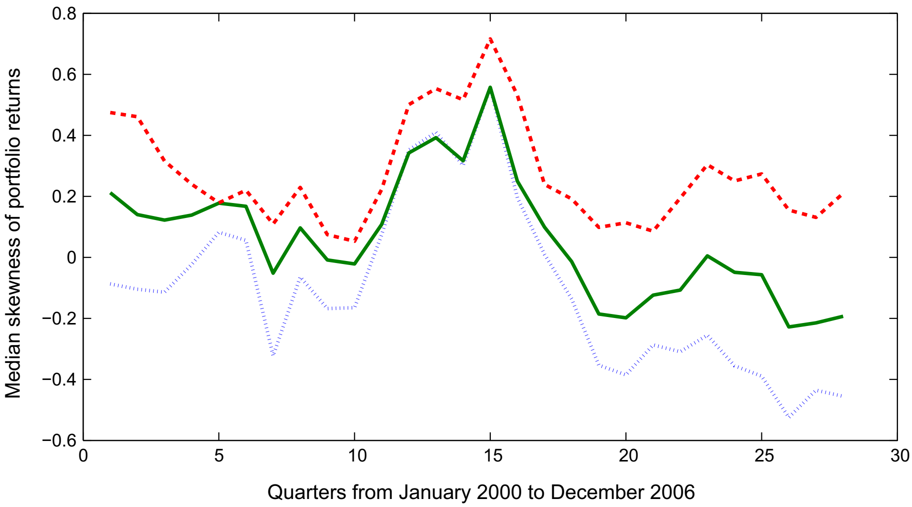

Fig. 4 Time-series of median skewness of portfolio returns. Evolution of median skewness of portfolio returns over the period 2000–2006 for deciles 1 (high diversification identified by a *dotted line*), 5 (*bold line*) and 10 (low diversification, *dashed line*)

These graphs confirm that the difference in skewness between decile 10 (low diversification) and decile 1 (high diversification) is almost always positive, but this difference is much higher in bull market periods, namely in the 5 first quarters and after quarter 20, i.e., from the middle of 2003 to the end of 2006. This result would align with our simulation results 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). The next section is devoted to this analysis. It also stresses the influence of portfolio value on diversification choices.

5 Empirical results

In this section, we show 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 by the kind of stocks composing the portfolio. Contrary to Mitton and Vorkink (2007), we find that the strength of the relationship depends on the sharing of global variance between systematic and idiosyncratic variance.

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

5.1 Skewness, diversification and common factors

We decompose the period 1999–2006 into 32 quarters, each of them being equivalent to a simulated market of Sect. 3. The number of trading days is approximately 65 per quarter; each trading day is "equivalent" to a state of nature. However, the number of stocks held by at least one investor is much greater than the number of days in a quarter. Consequently, in addition to the general test with all stocks, we perform supplementary tests with different numbers of stocks. We select stocks that are held by 0.5, 1 and 1.5 % of investors, respectively, 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 from the behavior of infrequently 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 6. 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 supplementary tests, this value 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 is highly significant, and lies between −0.724 in the one-factor model, and −0.774 in the three-factor model. However, even in the case with 76 stocks, the correlation is approximately −0.5, which is also highly significant.

These results confirm the link between the decrease in skewness due to diversification and the variance due to common factors. Moreover, for each quarter we calculate the average correlation coefficient of stock returns and the decrease of skewness due to diversification. Then we compute the rank correlation between the vector of average

Table 6 Rank correlations between the decrease in skewness due to diversification and the share of variance from the common factors

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***

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 $k = 1, \ldots, 5$. 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. As usual, *** denotes statistical significance at the 1 % level

correlations and the vector of decreases in skewness. This rank correlation varies from −0.31 to −0.59 across the four subsets of stocks in Table 6. In other words, there exists a strong link between skewness variation due to diversification choices and market conditions.

Choosing stocks held by a minimum percentage of investors biases the selection process toward diversified investors. In short, doing so could eliminate stocks that are prominently held by skewness seekers holding highly underdiversified portfolios. Our conclusion would be true, but the behavioral interpretations of this conclusion would be flawed. One way to address this problem is to duplicate the results of Table 6 with randomly chosen stocks. In each of the three panels A to C of Table 6, we denote n_i the number of stocks in quarter i. We draw at random $(n_1, \ldots, n_{32})$ stocks and perform the same calculations as in Table 6. We repeat 100 times the sequence, evaluating the significance of the rank correlation in each simulation. Table 7 summarizes the results. A given panel presents the number of cases in which 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 yield 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. The decrease in skewness is more significantly linked to diversification itself, than to the stock-picking skills of underdiversified investors.

Using a different methodology, Mitton and Vorkink (2007) found that investors holding underdiversified portfolios also pick highly skewed stocks. The abovementioned result does not support this conclusion. In our data, investors gamble at the portfolio level, not at the stock level.

Table 7 Percentages of significant rank correlations between the decrease in skewness due to diversification and the share of variance on the common factors

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

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

5.2 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 small number of stocks. According to Liu (2014), it is also optimal for "poor" investors to underdiversify if they are subject to solvency constraints. Liu shows that below a given amount invested in risky assets, the optimal portfolio contains only one stock. The optimal number of stocks thus increases with the amount to be invested in risky assets. Consequently, solvency constraints can also justify underdiversification for less wealthy investors.

However, Table 5 shows that less diversified portfolios normally generate a higher skewness in returns. This phenomenon also appeared in our theoretical results in Sect. 3.

To analyze the link between diversification, skewness seeking and solvency constraints in greater depth, we perform a multivariate analysis. We regress diversification measures on skewness and portfolio value in a panel data setting with fixed effects, as follows:[10]

$$D_{jt} = a_j + +a_P \ln(Portfolio\_Value_{jt}) + a_S Skewness_{jt} + \varepsilon_{jt} \tag{6}$$

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 t D_{jt} is measured, i.e., at the end of each semester or year, depending on the periodicity of the analysis.

[10] We significantly reject the hypothesis of a random effect model with the Hausman test at the highest level of significance in all models.

Table 8 Panel regression of diversification index on portfolio value and skewness of returns

	Half-yearly			Yearly		
	D_1	D_2	D_3	D_1	D_2	D_3
a_j	2.12***	1.5855***	1.4125***	2.0127***	1.5222***	1.3266***
	(172.24)	(290.32)	(252.66)	(147.55)	(244.77)	(198.41)
a_P	−0.3973***	−0.1284***	−0.0994***	−0.3871***	−0.1214***	−0.0885***
	(−281.89)	(−205.32)	(−155.43)	(−246.48)	(−169.60)	(−115.06)
a_S	0.0056***	0.0064***	0.0099***	0.0073***	0.0084***	0.0127***
	(8.70)	(20.88)	(21.69)	(7.32)	(17.25)	(17.39)
F	40,038.04	21,881.56	12,763.93	30,450.16	14,754.94	6,969.00
R^2 within	0.4430	0.2870	0.1249	0.4477	0.2715	0.0991
R^2 between	0.5913	0.4789	0.4303	0.5972	0.4779	0.4088
R^2 overall	0.5689	0.4082	0.2973	0.5783	0.4170	0.2948
Rho	0.7371	0.4082	0.5135	0.7585	0.6900	0.5234
Number of observations	779,907	779,907	779,907	358,502	358,502	358,502
Number of groups	80,671	80,671	80,671	76,825	76,825	76,825

The six columns correspond to three diversification indices (D_1, D_2 and D_3) and two ways to calculate the skewness of returns (with a semester, or a year of daily data). The regression equation is the following:

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

a_j denotes the intercept, a_P is the coefficient of portfolio value and a_S is the coefficient of skewness. The number of observations is the sum of the numbers of investors over the periods. Number of groups is 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. Standard deviations are adjusted for clustering and t-stats are given in parentheses. As usual, *** denotes statistical significance at the 1 % level

Panel data analysis allows us to control for individual heterogeneity. Each investor has his, or her, (unknown) own individual characteristics that may impact skewness, portfolio value or diversification (for instance, being a male or a female may influence the number of assets in portfolio). Fixed effects remove the impact of the time-invariant individual characteristics from the predictor variables.

In Table 8, we present semi-annual and annual results over the period 1999–2006. For semi-annual results, the skewness of returns is estimated over 6 months of daily returns and the panel data analysis is conducted over 16 semesters, whereas annual results are obtained over 7 years.

For each period, the three first columns refer to the three diversification measures D_1, D_2 and D_3. *N obs.* and *N groups* refer to the number of rows and the number of individual investors (or individual portfolios). Because our panel is unbalanced, individual investors are present 9.7 semesters and 4.7 years on average. In all analyses, the results are significant at the highest level.[11]

The first three lines of Table 8 provide the regression coefficients of Eq. (6) with t-stats shown in parentheses. For example, consider index D_1 in the semi-annual analysis. There are $80,671$ portfolios, $F = 40,038.04$ and the $\overline{R}^2$ coefficients are equal to 0.4430 (within), 0.5913 (between) and 0.5689 (overall). The intercept is 2.12, the coefficient of $ln(Portfolio\_Value)$ is -0.3973 and the coefficient of portfolio skewness is 0.0056.

The main comments are identical for the two analyses. First, all three coefficients are significant at the 1 % level. The most significant variable is clearly $ln(Portfolio\_Value)$ with t-stats varying from -115.06 to -281.89. 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 (2014) model because less-wealthy investors subject to solvency constraints hold a portfolio containing a very small number of stocks. As with D_1, estimates are -0.3871 with yearly data and -0.3973 with half-yearly data. We can conclude that increasing the portfolio value by 10 % leads to an average increase of 3.8 % in the number of stocks held by investors. This point was not considered by Mitton and Vorkink (2007). Moreover, using our estimates for D_1 with yearly values, we find that the average portfolio value of any investor holding only one stock is 181€, 1,087€ for two stocks, 11,599€ for 5 stocks, 69,532€ for 10 stocks or 1,188,209€ for 30 stocks.

Second, skewness is always significant, for D_1, D_2 or D_3, and the regression coefficients are positive. Considering different time windows to estimate skewness does not change the results, as shown in Table 8.

This stability means that, even when considering portfolio value, skewness remains significant (with the expected sign) in explaining diversification choices.[12]

However, diversification may also be explained by market trends but market trends impact skewness. For example, in bear markets, investors may decrease their number of stocks in portfolios but skewness of returns may also decrease simply because there are fewer positive returns to create skewness . To explore the intertemporal relationship

[11] Three different measures of the reliability of our results are provided: the F-statistic (testing whether the vector of regression coefficients is the null vector), the R^2s, overall, between and within, and finally the intraclass correlation *rho*, which is the fraction of the variance that is due to differences across individuals. Moreover, because there are multiple observations for each investor, the standard deviations of estimates are clustered at the investor level.

[12] It should be noted that we obtain the same results (unreported) when we exclude investors who hold only one stock in any sub-period. Because there are many observations that are clustered approximately at 1 for D_1, these systematically underdiversified investors do not drive our main conclusions.

Table 9 Panel regression of diversification index on portfolio value, skewness of returns and interaction variables for skewness and semesters

	Half-yearly					
	D_1	t-stat	D_2	t-stat	D_3	t-stat
a_j	2.1805***	175.70	1.6031***	291.52	1.4242***	252.73
a_P	−0.4034***	−284.35	−0.1300***	−206.79	−0.1008***	−156.44
1999$S1 \times Sk$	0.1089***	20.16	0.0373***	14.94	0.0364***	11.68
1999$S2 \times Sk$	0.1088***	31.15	0.0179***	11.05	0.0093***	4.43
2000$S1 \times Sk$	0.0545***	15.00	0.0176***	10.55	0.0326***	13.89
2000$S2 \times Sk$	0.0391***	11.01	0.0152***	8.76	0.0446***	18.17
2001$S1 \times Sk$	0.0089***	4.38	0.0032***	2.86	0.0076***	4.87
2001$S2 \times Sk$	−0.0533***	−19.30	−0.0213***	−15.64	0.0202***	11.48
2002$S1 \times Sk$	−0.0474***	−18.89	−0.0031***	−2.68	−0.0665***	−3.86
2002$S2 \times Sk$	−0.1644***	−68.59	−0.0455***	−47.62	−0.0273***	−23.97
2003$S1 \times Sk$	−0.0538***	−21.37	−0.0159***	−13.08	0.0216***	11.35
2003$S2 \times Sk$	−0.0360***	−18.16	−0.0067***	−6.99	0.0028*	1.77
2004$S1 \times Sk$	0.0451***	24.54	0.0185***	20.16	0.0115***	7.97
2004$S2 \times Sk$	0.0219***	12.66	0.0137***	15.44	0.0076***	5.23
2005$S1 \times Sk$	0.0428***	22.99	0.0235***	25.94	0.0295***	22.08
2005$S2 \times Sk$	0.0188***	9.81	0.0154***	16.55	0.0089***	7.01
2006$S1 \times Sk$	0.0028	0.82	0.0111***	6.91	−0.0469***	−17.75
2006$S2 \times Sk$	0.0858***	31.97	0.0354***	27.72	0.0330***	16.90
F	4,944.41		2,720.10		1,598.94	
R^2 within	0.4557		0.2944		0.1298	
R^2 between	0.5948		0.4811		0.4313	
R^2 overall	0.5734		0.4114		0.3000	
Rho	0.7398		0.6752		0.5142	
Number of observations	779,907		779,907		779,907	
Number of groups	80,671		80,671		80,671	

Columns 2 and 3 correspond to diversification index D_1, columns 4 and 5 correspond to D_2 (with associated t-statistics) and columns 6 and 7 correspond to D_3 (with associated t-statistics) . The regression equation is the following:

$$D_{jt} = a_j + a_P ln(PortfolioValue_{jt}) + a_1 1999S_1 \times Sk + a_2 2000S_2 \times Sk + \cdots + \varepsilon_{jt}$$

a_j denotes the intercept, a_P is the coefficient of portfolio value. The variables $Year_t S1 \times Sk$ and $Year_t S2 \times Sk$ are interaction variables for skewness and semesters (1 and 2 for each year). Number of observations is the sum of the numbers of investors over the periods. Number of groups is 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

between diversification and skewness, we add time-skewness interaction variables to Eq. (6). In Tables 9 and 10, we present semi-annual and annual results over the period 1999–2006 with time-skewness interaction variables.

Because the coefficients for a_j, and a_P remain stable, we note that only the coefficients of interaction variables for skewness reflect market trends only.

More precisely, in the two tables, the impact of skewness is significantly negative over 2002 for D_1, D_2 and D_3, and also over 2001 semester 2 and over 2003 for D_1 and D_2. These

Table 10 Panel regression of diversification index on portfolio value, skewness of returns and interaction variables for skewness and years

	Half-yearly					
	D_1	t-stat	D_2	t-stat	D_3	t-stat
a_j	2.0367***	148.89	1.5296***	245.47	1.3277***	197.88
a_P	−0.3895***	−247.36	−0.1221***	−170.22	−0.0887***	−114.91
2000 × *Sk*	0.0687***	16.02	0.0206***	10.61	0.0296***	11.35
2001 × *Sk*	0.0152***	6.62	0.0028***	2.27	0.0053***	3.08
2002 × *Sk*	−0.0389***	−15.16	−0.0006	−0.49	−0.0063***	−3.40
2003 × *Sk*	−0.0449***	−17.37	−0.0139***	−10.76	0.0204***	10.10
2004 × *Sk*	0.0274***	14.37	0.0146***	14.85	0.0157***	10.36
2005 × *Sk*	0.0337***	16.54	0.0228***	23.03	0.0391***	26.89
2006 × *Sk*	−0.0220***	−5.50	0.0018***	4.04	−0.0433***	−15.13
F	7,715.56		3,787.25		1,872.26	
R^2 within	0.4508		0.2736		0.1044	
R^2 between	0.5979		0.4781		0.4077	
R^2 overall	0.5794		0.4178		0.2966	
Rho	0.7590		0.6904		0.5251	
Number of observations	358,502		358,502		358,502	
Number of groups	76,825		76,825		76,825	

Columns 2 and 3 correspond to diversification index D_1, columns 4 and 5 correspond to D_2 (with associated t-statistics) and columns 6 and 7 correspond to D_3 (with associated t-statistics) . The regression equation is the following:

$$D_{jt} = a_j + a_P ln(PortfolioValue_{jt}) + a_1 2001 \times Sk + a_2 2002 \times Sk + \cdots + \varepsilon_{jt}$$

a_j denotes the intercept, a_P is the coefficient of portfolio value. The variables $Year \times Sk$ are interaction variables for skewness and years. The number of observations is the sum of the numbers of investors over the periods. Number of groups is 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

results suggest that during these bear markets, portfolio diversification is increasing with skewness, when portfolio value is taken into account (which also decreases in bear markets).

Finally, to take into account diversification inertia between sub-periods, we examine the dynamics of the relationship between diversification and skewness. We use the following level equation of the dynamic panel data model:

$$D_{jt} = \alpha D_{jt-1} + a_S Skewness_{jt} + a_P \ln(Portfolio\_Value_{jt}) + \sum_p \theta_p 1_p + \delta_j + \varepsilon_{jt} \quad (7)$$

where D_{jt-1} is the diversification index of the preceding sub-period and 1_p are time dummies[13] for sub-periods p , δ_j is an individual unobserved fixed effect and ε_{jt} is the error term, the other two variables having already been defined in the description of Eq. (6).

[13] We only present our results for years 2000 to 2006. We also estimate the model in which sub-periods are semesters, but due to a high number of instruments used, our Stata program does not run the complete estimation. Results including a constant instead of semester dummies are available upon request.

Table 11 Dynamic panel regression of diversification index on portfolio value and skewness of returns

	Yearly		
	D_1	D_2	D_3
LD	0.4374***	0.3878***	0.2983***
	(65.24)	(57.43)	(57.89)
a_P	−0.4214***	−0.1311***	−0.1070***
	(−170.93)	(−106.42)	(−74.94)
a_S	0.0240***	0.0132***	0.0300***
	(20.12)	(21.22)	(29.18)
2001	(dropped)	−0.0436***	(dropped)
2002	−0.1078***	−0.0635***	−0.0208***
2003	−0.1558***	−0.0896***	−0.0000
2004	−0.0408***	−0.0482***	0.0484***
2005	0.0271***	−0.0207***	0.0336***
2006	0.0921***	(dropped)	0.1657***
Wald Chi2	53,067.18***	28,738.23***	28,803.23***
Number of instrumental variables	22	22	22
Number of observations	210,471	210,471	210,471
Number of groups	57,596	57,596	57,596

The three columns correspond to three diversification indices (D_1, D_2 and D_3) and skewness of returns is computed over a year of daily data. *LD* is the coefficient of the lagged diversification index, a_P is the coefficient of portfolio value, a_S is the coefficient of skewness and 2001–2006 are the coefficients for yearly dummies. z-stats are given in parentheses. Wald Chi2 is the Chi2 stat with 8 degrees of freedom. Number of groups is the number of different investors entering the analysis. Two different investors may stay in the database for different time-lengths. As usual, *** denotes statistical significance at the 1 % level

In this specification, estimation by GMM for dynamic panel data is the only unbiased procedure, which is called the Arellano–Bond estimation (Arellano and Bond 1991) in the case of GMM in differences.[14] In the Arellano–Bond methodology, Eq. (7) is estimated by taking the first differences of variables:

$$\Delta D_{jt} = \alpha \Delta D_{jt-1} + a_S \Delta Skewness_{jt} + a_P \Delta \ln(Portfolio\_Value_{jt}) + \Delta\varepsilon_{jt} \quad (8)$$

This differentiation process removes the unobserved fixed effects. As many instrumental variables are used (all lagged independent variables are valid instruments), the model is over-identified which makes necessary the use of a GMM estimator and a joint test of model specification and validity of the instruments (Sargan test of over-identifying restrictions is $chi2(14) = 616.6615$).

In Table 11, we present annual results for the period 2001–2006 for the three diversification indexes.[15] Results show that coefficient estimates are significant at the highest

[14] The presence of a lagged variable with fixed effects produces biased and inconsistent OLS estimates, which occurs because the lagged dependent variable is correlated with the error term although there is no autocorrelation between terms ε_{jt}.

[15] Due to variable differentiation, only 57,596 (versus 76,825 in Table 8) investors are examined over a maximum length of 6 years. The estimation uses a total of 22 instrumental variables. Due to the huge number of instrumental variables, we do not perform the same analysis over semesters.

level. More importantly, we observe highly significant $\alpha < 1$, which indicates a decrease in all diversification indexes over time. Once again, estimates for skewness are positive. Therefore, the third moment of returns increases diversification changes. In the same vein, market trends and investors' portfolio value still decrease diversification changes over time. In other words, we do not detect any variation in the relationship between diversification and skewness over time when we control for investor portfolio value and market trends. It should be noted that these results also hold when we exclude single-stock holders. For example, $\alpha = 0.4362$ with D_1 for the remaining 45,911 single-stock holders.

6 Concluding remarks

Building highly skewed portfolios may reveal either a propensity to gamble or prudent behavior. In this paper, we first demonstrate analytically that diversification in Arrow–Debreu markets decreases variance, but also the skewness of returns. We also show that this relationship between diversification and skewness holds 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 and confirm some of the empirical results obtained by Mitton and Vorkink (2007) for 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 our theoretical results. However, when determining whether undiversified investors concentrate on highly skewed stocks, we obtain ambiguous results. More precisely, 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 explore this relationship in greater depth, we show the negative relationship between the decrease in skewness due to diversification and the share of variance explained by a given number of common factors.In other words, in bearish markets characterized by a strong market factor and a high average correlation between stock returns, underdiversification and skewness in portfolio returns are not significantly related.

Finally, we test whether skewness remains significant in the explanation of diversification choices after controlling for portfolio value. For less wealthy investors, managing a diversified portfolio is costly; they may find optimal to hold a few stocks. Our panel data regression shows that portfolio value is the main determinant of diversification choices. Nevertheless, skewness of returns remains significant after controlling for portfolio value.

The contribution of our paper goes beyond Mitton and Vorkink (2007) in several directions. We first establish a theoretical link between diversification and skewness, and second, we show that the strength of this relationship depends on the "concentration" in the market, i.e., the sharing of global variance between idiosyncratic and systematic variance. We were able to analyze this aspect because of our database of individual investors covering an eight-year period, including the dotcom bubble burst and the subsequent partial recovery.

Acknowledgments We thank an anonymous referee for suggesting improvements in preceding versions of the paper. We also thank Laurent Deville, Gunter Franke, Burton Hollifield, Jens Jackwerth, Gregory Nini, Charles Noussair, Winfried Pohlmeier, Mark Seasholes, Pierre Six, Marc Willinger, the participants of the DMM meeting (2011, Montpellier), the Konztanz-Strasbourg Workshop (2011, Königsfeld), the Behavioral Insurance Meeting (2011, München), the French Finance Association Meeting (2014) for comments and suggestions. The financial supports of OEE (Observatoire de l'Epargne Européenne) and CCR Asset Management are gratefully acknowledged. We thank Tristan Roger for valuable research assistance and computer programming.

Appendix: Moments of equally-weighted portfolios of Arrow–debreu securities

Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ denote a finite state-space with n equally-likely states of nature and assume that all Arrow–Debreu securities, denoted $X_1, \ldots, X_n$, are traded. X_i pays 1 in state ω_i and 0 elsewhere. $(p_1, \ldots, p_n)$ stands for a sequence of equally-weighted portfolios containing respectively $1, 2, \ldots, n$, AD securities. Without loss of generality, we assume that p_k contains $1/k$ units of each of the first k securities.[16]

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

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

$$E(X_k) = m_k = \frac{1}{n} \tag{9}$$

$$V(X_k) = \frac{n-1}{n^2} \tag{10}$$

$$E\left[(X_k - m_k)^3\right] = \frac{(n-1)(n-2)}{n^3} \tag{11}$$

$$cov(X_k, X_{k^*}) = -\frac{1}{n^2} \text{if} \mathrm{k} \neq \mathrm{k}^* \tag{12}$$

Proof 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

$$\begin{aligned} E\left[(X_i - \mu_i)^3\right] &= E(X_i^3) - 3\mu_i E(X_i^2) + 3\mu_i^2 E(X_i) - \mu_i^3 \\ &= \frac{1}{n} - 3\frac{1}{n^2} + 3\frac{1}{n^3} - \frac{1}{n^3} = \frac{1}{n} - \frac{3}{n^2} + \frac{2}{n^3} \\ &= \frac{(n-1)(n-2)}{n^3} \end{aligned} \tag{13}$$

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 $i \neq j$.

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.

Proposition 2

$$\begin{aligned} \forall k, \mu_k &= \frac{1}{n} \\ \sigma_k^2 &= \frac{1}{n}\left(\frac{1}{k} - \frac{1}{n}\right) \end{aligned} \tag{14}$$

Proof Proposition 1 allows to write the covariance matrix of the n AD securities payoffs as

[16] As 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.

$$\mathbf{V}_n = \frac{1}{n}\mathbf{I}_n - \frac{1}{n^2}\mathbf{1}_{(n,n)} \tag{15}$$

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

$$\sigma_k^2 = \frac{1}{k^2}\mathbf{1}'_{(k)}\mathbf{V}_k\mathbf{1}_{(k)} \tag{16}$$

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

$$\begin{aligned}\sigma_k^2 &= \frac{1}{k^2}\mathbf{1}'_{(k)}\mathbf{V}_k\mathbf{1}_{(k)} = \frac{1}{k^2}\mathbf{1}'_{(k)}\left(\frac{1}{n}\mathbf{I}_k - \frac{1}{n^2}\mathbf{1}_{(k,k)}\right)\mathbf{1}_{(k)} \\ &= \frac{1}{k^2 n}\mathbf{1}'_{(k)}\mathbf{I}_k\mathbf{1}_{(k)} - \frac{1}{k^2 n^2}\mathbf{1}'_{(k)}\mathbf{1}_{(k,k)}\mathbf{1}_{(k)} \\ &= \frac{1}{kn} - \frac{1}{n^2} = \frac{1}{n}\left(\frac{1}{k} - \frac{1}{n}\right)\end{aligned} \tag{17}$$

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.

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 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:

$$s_k^3 = E\left(\left(\frac{1}{k}\sum_{j=1}^{k} X_j - \mu_k\right)^3\right) = \frac{1}{k^3}E\left(\left(\sum_{j=1}^{k} X_j - \frac{k}{n}\right)^3\right) \tag{18}$$

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:*

$$s_k^3 = \frac{1}{n^3}\left[\left(\frac{n}{k} - 1\right)\left(\frac{n}{k} - 2\right)\right] \tag{19}$$

Proof

$$\begin{aligned}s_k^3 &= \frac{1}{k^3}E\left[(Y_k^3 - \left(\frac{k}{n}\right)^3 - 3\left(\frac{k}{n}\right)Y_k^2 + 3\left(\frac{k}{n}\right)^2 Y_k\right] \\ &= \frac{1}{k^3}\left[\frac{k}{n} - \left(\frac{k}{n}\right)^3 - 3\left(\frac{k}{n}\right)^2 + 3\left(\frac{k}{n}\right)^3\right]\end{aligned} \tag{20}$$

Rearranging terms leads to

$$\begin{aligned} s_k^3 &= \frac{1}{n^3}\left[\left(\frac{n}{k}\right)^2 - 3\left(\frac{n}{k}\right) + 2\right] \\ &= \frac{1}{n^3}\left[\left(\frac{n}{k} - 1\right)\left(\frac{n}{k} - 2\right)\right] \end{aligned} \tag{21}$$

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:

$$s_k^3 = \frac{1}{n}\left[\left(D_1 - \frac{1}{n}\right)\left(D_1 - \frac{2}{n}\right)\right] \tag{22}$$

We can also establish a very simple relationship between s_k^3 and σ_k^2 using Eqs. (14) and (19).

$$s_k^3 = \sigma_k^2\left(D_1 - \frac{2}{n}\right) \tag{23}$$

References

Albuquerque R (2012) Skewness in stock returns: reconciling the evidence on firm versus aggregate returns. Rev Financ Stud 25:1630–1673

Ang A, Bekaert G (2002) International asset allocation with regime shifts. Rev Financ Stud 15:1137–1187

Ang A, Bekaert G (2004) How do regimes affect asset allocation? Financ Anal J 60:86–99

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. Rev Econ Stud 58:277–297

Baker M, Wurgler J, Yuan Y (2012) Global, local, and contagious investor sentiment. J Financ Econ 104:272–287

Bali TG, Cakici N, Whitelaw RF (2011) Maxing out: stocks as lotteries and the cross-section of expected returns. J Financ Econ 99:427–446

Barber B, Odean T (2000) Trading is hazardous to your wealth: the common stock investment performance of individual investors. J Finance 55(2):773–806

Barberis N, Huang M (2008) Stocks as lotteries: the implications of probability weighting for security prices. Am Econ Rev 98:2066–2100

Blume ME, Friend I (1975) The asset structure of individual portfolios and some implications for utility functions. In: Rodney L. White Center for Financial Research Working Papers 10–74. Wharton School Rodney L. White Center for Financial Research

Brunnermeier MK, Gollier C, Parker JA (2007) Optimal beliefs, asset prices, and the preference for skewed returns. NBER Working Papers 12940. National Bureau of Economic Research Inc

Brunnermeier MK, Parker JA (2005) Optimal expectations. Am Econ Rev 95:1092–1118

Calvet LE, Campbell JY, Sodini P (2007) Down or out: assessing the welfare costs of household investment mistakes. J Polit Econ 115:707–747

Campbell R, Koedijk K, Kofman P (2002) Increased correlation in bear markets. Open Access publications from Maastricht University urn:nbn:nl:ui:27–19571. Maastricht University

Chamberlain G, Rothschild M (1983) Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica 51:1281–1304

Chen J, Hong H, Stein JC (2001) Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices. J Financ Econ 61:345–381

Conrad J, Dittmar RJ, Ghysels E (2013) Ex ante skewness and expected stock returns. J Finance 68:85–124

Cook PJ, Clotfelter CT (1993) The peculiar scale economies of lotto. Am Econ Rev 83:634–643

Deck C, Schlesinger H (2010) Exploring higher order risk effects. Rev Econ Stud 77:1403–1420
Ebert S, Wiesen D (2011) Testing for prudence and skewness seeking. Manag Sci 57:1334–1349
Eeckhoudt L, Schlesinger H (2006) Putting risk in its proper place. Am Econ Rev 96:280–289
Entrop O, McKenzie M, Wilkens M, Winkler C (2014) The performance of individual investors in structured products. Rev Quant Finance Account. doi:10.1007/s11156-014-0479-8
Forrest D, Simmons R, Chesters N (2002) Buying a dream: alternative models of demand for lotto. Econ Inq 40:485–496
Foucault T, Sraer D, Thesmar DJ (2011) Individual investors and volatility. J Finance 66:1369–1406
Francis JC, Kim D (2013) Modern portfolio theory: foundations, analysis and new developments. Wiley, New York
Garrett TA, Sobel RS (1999) Gamblers favor skewness, not risk: further evidence from united states' lottery games. Econ Lett 63:85–90
Goetzmann WN, Kumar A (2007) Equity portfolio diversification. Rev Finance 12:433–463
Gollier C (2005) Optimal illusions and decisions under risk. CESifo Working Paper Series 1382. CESifo Group Munich
Harding MC (2008) Explaining the single factor bias of arbitrage pricing models in finite samples. Econ Lett 99:85–88
Harvey CR, Siddique A (1999) Autoregressive conditional skewness. J Financ Quant Anal 34:465–487
Harvey CR, Siddique A (2000) Conditional skewness in asset pricing tests. J Finance 55:1263–1295
Karolyi GA, Stulz RM (2003) Are financial assets priced locally or globally? In: Constantinides G, Harris M, Stulz RM (eds) Handbook of the economics of finance, volume 1 of Handbook of the Economics of Finance. chapter 16. Elsevier, Amsterdam, pp 975–1020
Kelly M (1995) All their eggs in one basket: portfolio diversification of us households. J Econ Behav Organ 27:87–96
Kim WC, Fabozzi FJ, Cheridito P, Fox C (2014) Controlling portfolio skewness and kurtosis without directly optimizing third and fourth moments. Econ Lett 122:154–158
Kimball MS (1990) Precautionary saving in the small and in the large. Econometrica 58:53–73
Kraus A, Litzenberger RH (1976) Skewness preference and the valuation of risk assets. J Finance 31:1085–1100
Kumar A (2007) Do the diversification choices of individual investors influence stock returns? J Financ Mark 10:362–390
Kumar A (2009) Who gambles in the stock market? J Finance 64:1889–1933
Lease RC, Lewellen WG, Schlarbaum GG (1974) The individual investor: attributes and attitudes. J Finance 29:413–433
Lesmond DA, Ogden JP, Trzcinka CA (1999) A new estimate of transaction costs. Rev Financ Stud 12:1113–1141
Lewis KK (1999) Trying to explain home bias in equities and consumption. J Econ Lit 37:571–608
Liu H (2014) Solvency constraint, underdiversification, and idiosyncratic risks. J Financ Quant Anal 49:1–39
Longin F, Solnik B (2001) Correlation structure of international equity markets during extremely volatile periods. J Finance 56:649–676
Menezes C, Geiss C, Tressler J (1980) Increasing downside risk. Am Econ Rev 70:921–932
Mitton T, Vorkink K (2007) Equilibrium underdiversification and the preference for skewness. Rev Financ Stud 20:1255–1288
Nagel S (2005) Short sales, institutional investors and the cross-section of stock returns. J Financ Econ 78:277–309
Odean T (1999) Do investors trade too much? Am Econ Rev 89:1279–1298
Ortobelli S, Biglova A, Huber I, Racheva-Iotova B, Stoyanov S (2005) Portfolio choice with heavy-tailed distributions. J Concr Appl Math 3:353–376
Shefrin H, Statman M (2000) Behavioral portfolio theory. J Financ Quant Anal 35:127–151
Shu PG, Chiu SB, Chen HC, Yeh YH (2004) Does trading improve individual investor performance? Rev Quant Finance Account 22:199–217
Stoyanov S, Rachev S, Racheva-Iotova B, Fabozzi F (2011) Fat-tailed models for risk estimation. J Portf Manag 37:107–117
Tarazona-Gomez M (2004) Are individuals prudent? An experimental approach using lottery choices. Technical Report, Copenhagen Business School
Tversky A, Kahneman D (1992) Advances in prospect theory: cumulative representation of uncertainty. J Risk Uncertain 5:297–323
Walker I, Young J (2001) An economist's guide to lottery design. Econ J 111:700–722
Xu J (2007) Price convexity and skewness. J Finance 62:2521–2552