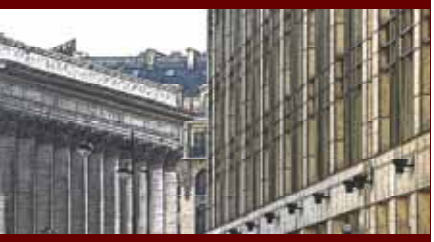
March 2007

Research Department



Working Papers - n°3

Can today's stars be used to read the stars of the future?

Case study of equity funds distributed in France between 2000 and 2005

Olivier Garnier* et Thierry Pujol**[1]

(*) Member of the Scientific Advisory Board of the Autorité des Marchés Financiers (AMF) and Deputy General Manager of Sociéte Générale Asset Management

(**) Partner of DPA Invest

[1] The authors thank *Morningstar* and *Standard & Poor's* for making their fund databases available. Of course, the authors retain full responsibility for this study and the two rating agencies are in no way liable for the views expressed herein

AUTORITÉ DES MARCHÉS FINANCIERS

AMF

Abstract

Growth in the distribution of funds through external channels has lead to increased recourse to ratings. This practice, already used in Anglo-American markets for a long time, has recently gained ground in Continental Europe, particularly in France. The most widely ratings available to individual investors, such as Morningstar or Standard & Poor's star ratings, are based on risk-adjusted measurements of past returns. As such, they are not designed in principle as tools to predict future performance. However, given the potential of using such ratings for this purpose, we have empirically tested their persistence over time. In other words, what is the probability that funds with star ratings today retain their ratings in the future and over what span of time?

To answer this question for equity funds (French and euro area) distributed in France, we used two data sets made available to us by Morningstar and Standard & Poor's, including funds rated monthly by these two agencies over 2000-2005. Our analysis based on a probabilistic method for the estimation of transition matrices has produced two observations. Firstly, there is indeed a certain degree of persistence for star ratings, though low and, above all, it tends to disappear rapidly beyond a three-year time span (the historical period for their calculation). Some cases can nevertheless be identified where persistence is higher. On the one hand, funds with consistently high ratings in the past ("seasoned stars") have a greater chance of remaining highly rated in the future (compared to recently upgraded funds). On the other hand, funds with initial high ratings (the "young stars"), recently rated (after passing the historical three-year milestone) and included directly in the top-rated category, have also higher chances of retaining their position. The second observation concerns the surprising variation among star-ratings from one month or quarter to another. This is explained by the visible impact of movements into and out of the ratings classifications, with newly rated funds tending to be better rated on average, while the contrary is true for funds exiting the ratings. This rotation thus contributes to downgrading funds already 'starred'. The selection bias appears more pronounced on entry than on exit. This phenomenon is interesting because up to date most studies have focused on the impact of the survivorship bias, neglecting the "incubation" bias.

This study provides the basis for two simple recommendations to improve the robustness of information provided to investors by ratings: increase transparency on the entry into and exiting from fund ratings and provide historical data for previously 'starred' funds and not only data for those of the last available month.

1 - Introduction

With the development of open architecture for the distribution of collective investment funds, recourse to rating systems has increased. This practice, already prevalent for some time in Anglo-American markets, has recently made inroads in Continental Europe, particularly in France.

Ratings most commonly available to individual investors are those of Morningstar and Standard & Poor's based on measurement of past returns. However, to quote the widely used formulation, past performances do not guarantee future results. In consequence, agencies using such rating mechanisms take considerable care by issuing very explicit disclaimers. On the one hand, such ratings must not be used to forecast future performances or select the "best" fund of a given category. Rather they should only be used to provide a robust basis to compare past results of similar funds, taking into account the risks. On the other hand, these ratings constitute only one source of information among many other. Funds should be selected on the basis of other criteria such as investment process and style, the stability of the fund management team, the asset allocation of the portfolio, etc. Moreover certain agencies moreover also publish qualitative ratings produced from in-depth analysis of these criteria on a fund by fund basis, notably through individual interviews.

Notwithstanding such disclaimers, investors tend to use these ratings assuming implicitly that their performances are relatively persistent. For example, in the US, according to the study by Del Guercio and Tkac (2001) over the period 1996-1999, a Morningstar 5-star rating results in an abnormal increase of 50% (on average) above the normally expected flow for six-month period. Their study also highlighted the significant impact on flows (in the expected direction) in response to upgrades and downgrades. It is likely that the commercial strategy of funds producers and distributors accentuates this phenomenon by encouraging them to more actively promote higher rated funds.

Given the use of these ratings, an empirical examination of their level of persistence is warranted. Does a fund with a 5-star or a 4-star rating today have a better chance than other funds to retain its ranking among top rated funds in the future? To this purpose, we have adopted the perspective of the individual investor, deliberately limiting the scope of questions raised as opposed to those addressed in more academic discussions on market efficiency and the pertinence of active management:

- We have focused on the persistence of the relative risk-adjusted performance of funds as compared with other funds, and not compared with benchmark indexes or a multifactor model. In other words, our goal is only to determine if better rated funds consistently outperform their peers in subsequent ratings and not whether they generate a consistent "alpha";
- We test only to see whether the ratings are persistent or not, without seeking to determine causes for possible persistence. There may be many such causes, including not only differences in the underlying skill of managers, but also management fees or investment styles.

The purpose of our study is not to identify the “best” rating system but rather to analyze the factors of persistence for the two rating systems adopted and used in the French mutual fund market over the longest period:

- Our analysis has been limited to the Morningstar and Standard & Poor’s rating mechanisms as they provide data for ratings for a relatively long period and are the most widely used. It should be noted that the most recent ratings calculated by Fundclass/Aptimum and EuroPerformance/Edhec Style rating explicitly incorporate persistence indicators in their criteria (though using different methods);
- We have not sought to compare methodologies underlying the different available rating systems. In particular, we do not address the frequently debated and complex question concerning the "appropriate" adjustment of the return in relation to risk incurred[2].

Finally it should be emphasized that persistence should not be the only criterion used to evaluate the pertinence of a rating:

- Persistence is a necessary condition but not sufficient for ratings to be used as a tool to select funds. For example, a lexicographical classification of mutual funds is by definition very stable though it does not constitute a meaningful source of information to determine relative future performance;
- Conversely, non-persistence of a rating should not be interpreted as an indication of a flaw in methodology. If mutual fund performances were completely arbitrary, any fund rating system would be completely unstable, however sophisticated.

In relation to existing literature on this subject, this study has several additional features:

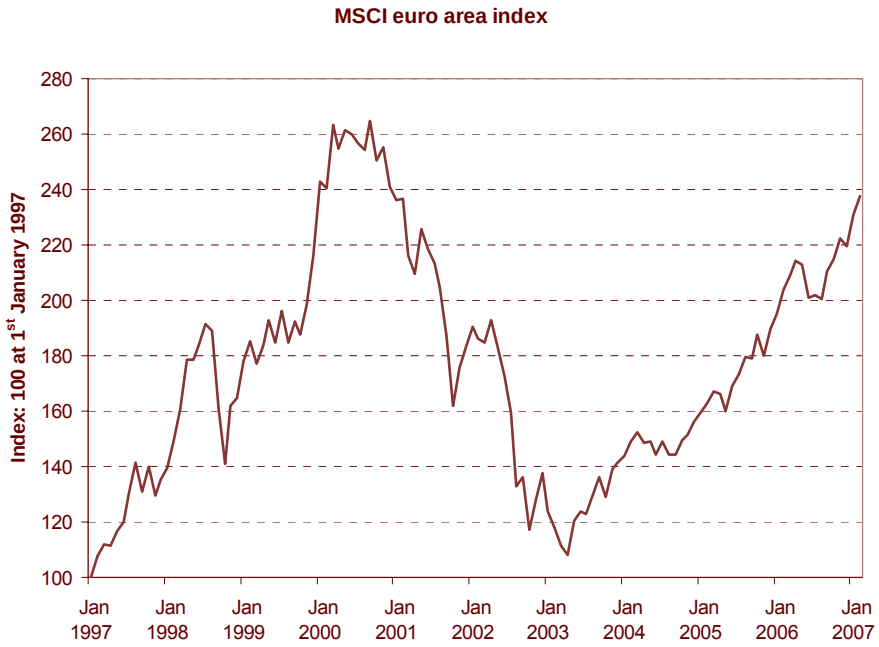
- It uses the ratings published monthly by Morningstar and Standard & Poor’s, whereas most other studies have established their rankings on the basis of panel data that is frequently partial or that fails to adequately take into account actual entry and exits of funds in the rankings;

[2] *For a comparative analysis of ratings, see Amenc-Le Sourd (2005).*

- It concerns UCITS equity funds (France and euro area) marketed in France, while most available studies focus on the US mutual funds market and, to a lesser extent, on the UK market;

- It covers 2000-2005 (thus reflecting the impact of the stock market bubble of the late 90s since the ranking for 2000 is based on performances of the three previous years), whereas the few studies covering the French market are based on data from the 1990s. The rankings established over 2000-2005 not only better reflect the recent situation of the fund management market but also several sub-periods marked by highly contrasting stock market trends (cf.. chart 1: highlighting the exceptional bull market cycle for stock market indices up to 2000, a sharp downturn until early 2003, followed by a net recovery from that time);

Chart 1: Equities markets performances in euro area



- Finally, it applies a stochastic approach based on estimations of performance matrices for the different star-rating levels. This makes it possible to draw on information provided by monthly rating changes and to obtain more accurate estimates than from data collected annually. It also takes into account movements of entries into and exits from ratings by adding in the transition matrix a category for "non-starred" funds.

- The outline of this study is as follows. After a brief overview of existing literature on the persistence of fund ratings (section II), section III describes the data providing the basis for our study and IV the method used. The main results are then presented and analyzed in section V. The last section offers a summary of key results.

2 - Overview of existing literature on ratings persistence

A rich and extensive body of literature has analyzed the US mutual funds market and, to a lesser extent, the UK market[3]. This research has offered the following insights:

- A significant survivorship bias exists : funds with lower performance are more likely to be closed ;
- However, after having adjusted for this bias, a certain level of performance persistence remains. This is generally greater when measured in relation to the peer group rather than stock market benchmark indexes ;
- Persistence is greater when analyzed over a shorter period, tending to disappear over a longer term ;
- The phenomenon of repetition is more pronounced for "losers" than for "winners" ;
- Finally, all these persistence factors do not only reflect the skill of managers but also other factors: management fees, style-level momentum or asset-level momentum, etc…

There are fewer studies on UCITS marketed in France. De Marchi (2006) produced a recent overview covering the period from 1993 to 2002. Existing literature concerns only "French equities" for which performance persistence is tracked only on an intermittent basis over the 1990s and at best only for short periods (year-on-year).

It should be emphasized that all the studies covering both Anglo-American and French markets draw up their respective rankings for funds on the basis of specific performance measurements and datasets. For this reason, they do not provide a means to directly determine the level of persistence for rankings published monthly by the main rating agencies and therefore available to individual investors. The few studies covering this specific issue concern the Morningstar ratings for the US market which offer highly contrasted results. The first major study by Khorana and Nelling (1998) demonstrated high persistence for Morningstar-rated funds between December 1992 and June 1995. Their conclusions were subsequently challenged by Blake and Morey (2000) showing that persistence results were overstated because of a survivorship bias. After correcting for this bias, 5-star funds did not appear to consistently outperform over the 1992-1997 period. Only the underperformance of 1-star or 2-star funds produces persistence results[4]. Morey (2003) goes even further, qualifying a 5-star rating as a "kiss of death". According to this latter study, during three years after

[3] *For a review of this literature, see for example Giles, Wilsdon and Worboys (2002).*

[4] *This study as well as those undertaken subsequently do not specifically address the persistence of Morningstar ratings, but rather their ability to predict future returns of funds. There may exist a degree of methodological inconsistency as future performances are measured according to a criteria not used by Morningstar. In our study we have decided to compare future with present ratings established according to the same criteria to prevent this bias.*

receiving its initial 5-star rating, fund performance severely falls off while risk levels increase. One possible interpretation for this phenomenon could be as a response to an upsurge in asset inflows after receiving a 5-star rating. This might make it difficult for the fund manager to pursue the same investment strategy as before (and notably unable to load on selected momentum stocks as before), encouraging greater risk-taking to maintain the ratings. However, a more recent study by Morey with Gottesman (2006) has radically questioned previous conclusions. This latter study looked at fund performance over the three-year period between June 2002 and June 2005, taking into account Morningstar's new methodology to rate funds introduced in June 2002[5]. For "US equity" funds, the Morey-Gottesman (2006) study concluded that the new Morningstar rating system can predict future performance for a three-year period. Moreover, the effect is relatively monotonic over the entire rating spectrum as even the next to lowest rated funds (2-star) significantly outperform the lowest (1-star). Studies published by Morningstar in the United States appear to support these conclusions. For example, according to Kinnel (2005, 2006), higher-rated funds in June 2002 or June 2003 outperformed other funds over the following three years, both in terms of the risk-adjusted performance used by Morningstar and absolute performance.

Given the evidence of persistence in the new Morningstar ratings for the US market, an examination of the French case for the most commonly used rating systems is warranted.

We have used two datasets for equity funds marketed in France provided to us respectively by Standard & Poor's (S&P) and Morningstar.

[5] These changes include notably the following: on the one hand, categories of funds distributed in the United States increased from 4 (domestic equity, international equity, municipal bonds and taxable bonds) to 48 to better reflect differences in investment styles (for example, for equities: "Large Cap" vs. "Small Cap"; "Value" vs. "Growth"); in addition the measure of performance used was modified to better factor in risk.

3 - The selection of UCITS and rating mechanisms used

3-1 The Standard & Poor's dataset

The dataset provided by S&P includes UCITS distributed in France (*Sicav*, *FCP* and foreign funds licensed by the AMF) for all or part of the December 2000 - December 2005 period belonging to one of two categories: "French equities" and "euro area equities" (according to the classification adopted by S&P).

For each month, we are provided with the stars assigned by S&P according to the methodology summarized in box 1. For funds included at the end of each year, we thus have data on their prior monthly returns.

Box 1: Standard & Poor's stars

S&P fund stars are calculated using the funds' monthly performance relative to its sector average over the three years to date. For the purpose of this study only UCITS distributed in France in each category are considered.

The following ratio is used to rank funds in relation to their peers (calculated over the preceding three years):

Average monthly relative performance/volatility of monthly relative performance

A denominator for the volatility of relative performance is included in order to take risks into account. This indicator represents an information ratio, calculated not in relation to the benchmark index for the fund in question but rather the average performance of funds of this category. Appendix 1 presents a comparative analysis between the rating obtained on the basis of this indicator and the rating that would have been obtained without adjusting for risk (i.e. based only on returns of the past three years).

On the basis of the ranking obtained from this indicator, the breakdown of stars assigned to funds would be as follows:

- 5-star: top 0-10%
- 4-star: 10-30%
- 3-star: 30-50%
- 2-star: 50-75%
- 1-star: 75-100%

It should however be noted that each fund included in this dataset is not necessarily rated every month over the December 2000 - December 2005 period. Indeed, a fund cannot be rated until it has a minimum period of historical data of 36 months in its respective category. As a result, there are many reasons why a fund is not rated for a given month:

- It was not yet created on the date in question;
- It has been created but it does not yet have a minimum record of historical data for 36 months;
- It has more than 36 months of existence but has changed category within less than 36 months[6];
- It no longer exists or has merged with another.

Also included in the dataset are funds that have never been rated such as those created after 2003 for which S&P does not possess sufficient historical data (36 months) to rank them.

As it will be shown below in greater detail, it is essential to have panel data limited exclusively to funds with star-ratings over an uninterrupted period during entire time frame under review. In this way it is possible to take into account potential bias in selection among funds added to or removed from the universe of rated funds. However, on the basis of the dataset provided by S&P (and by Morningstar), we made a few adjustments:

- We eliminated from our data panel funds rated for short periods (less than 6 months) before disappearing. These frequently represent funds transferred to other categories. Their inclusion would represent create "noise" unfairly increasing the volatility of rated funds;
- We also identified here and there interruptions in ratings of short duration (1 to 3 months) reflecting temporary reporting problems for maintaining the database. In such cases, we filled in the "gaps" on the basis of star-ratings assigned for the closest month.

The resulting data panel includes 513 funds with less than one half rated on an uninterrupted basis from the end of 2000 to the end of 2005.

[6] *In certain cases, a fund changing category can retain its historical data and consequently its star-ratings, if this change is not the result of a change in the fund's objectives: this could be the case, for example, if initially it was included in a category that was less appropriate or upon migration to newly created a category providing a better fit.*

3-2 The Morningstar dataset

The dataset provided by Morningstar is comparable in structure to S&P. We consequently applied to it the same adjustments described above. This provided us with panel data of 784 funds rated by Morningstar for at least 6 months during the period from October 2001 – September 2005[7].

The differences in rating methods used by Morningstar (cf. box 2) and S&P should be noted:

- The risk-adjusted performance indicator of Morningstar is obtained by means of a utility function with a constant relative risk aversion rather than a ratio of relative performance/relative volatility;
- Categories adopted for rating funds are not the same: for Morningstar, euro area equity funds are divided into three categories according to market capitalizations of the companies in which they invest (large, medium and small caps). In contrast to S&P, it does not include a category for French equities;
- For Morningstar, star-ratings are assigned by comparing (within each category) all funds distributed in Europe rather than only those distributed in France (as for S&P). As a result, for example, the percentage of 5-star funds out of all funds distributed in France (and rated by Morningstar) is not necessarily 10%;
- Stars are not assigned on the basis of the same breakdown: the median is in the middle of the 3-star range with Morningstar whereas for S&P it is located at the upper limit between 3-star and 4-star ratings.

[7] *The database provided by Morningstar initially covered a longer period (March 2001-March 2006). However we reduced the range of the dataset at both ends. On the one hand, between March and September 2001, an exceptionally large number of funds entered the ratings, which had a major impact on the hierarchy within this universe and thus the star ratings. On the other hand, Morningstar introduced new categories starting in November 2005, resulting in significant changes in the ratings.*

Box 2: Morningstar ratings

Morningstar ratings are calculated according to the MRAR (Morningstar Risk-Adjusted Return) method. The risk-adjusted performance indicator is determined by means of the utility function with a constant relative risk aversion. This is calculated as follows:

$$\mathrm{MRAR}(\gamma)=\left[\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}}\left(1+r_{Gt}\right)^{-\gamma}\right]^{-\frac{12}{\gamma}}-1$$

where r_{Gt} is the excess return (in relation to the return on a risk-free investment) and where γ describes the level of investor risk aversion.

Morningstar stars are assigned by classifying funds according to the value of this indicator MRAR over three years, where γ = 2 (with 2 considered by Morningstar as the typical investor risk aversion level).

On the basis of the ranking from this indicator, the breakdown of stars assigned to funds would be as follows:

- 5-star: top 0-10%
- 4-star: 10-32.5%
- 3-star: 32.5-67.5%
- 2-star: 67.5-90%
- 1-star: 90-100%

4 - Methodology

To analyze the level of persistence of fund stars, we adopted a probabilistic method that involves estimations of transition matrices among the different star rankings. Based on our review of existing literature, this method, which is widely used for credit ratings of issuers (to evaluate the probability of rating changes), has surprisingly not yet been applied for the rating of funds. In this way, we will demonstrate that it may contribute to an optimal analysis of monthly information available on trends for rated funds.

4-1 The limitations of analyses based on low frequency observations

To highlight the persistence of ratings of US mutual funds, Morningstar[8] analysts underscore the following results for the recent period: funds initially rated 5-stars (in June 2002 or June 2003) are rated slightly higher three years later than funds which initially had a lower rating. For example, in the "US equities" category, funds initially rated 5-stars in June 2003 have an average rating of 3.2 in June 2006 versus 3.0 for funds initially rated 4-stars, 2.8 for funds initially rated 3-stars and 2.7 for those initially rated 2-stars or 1-star.

The same type of descriptive analysis can be conducted on the two datasets available to us for equity funds distributed in France. The results are presented below in charts 2 (Morningstar) and 3 (S&P). In both cases funds are divided into five groups according to their initial ratings (September 2001 or December 2000). For each group the average number of stars obtained for the following years by rated funds is then calculated. Overall, this analysis highlights rapid convergence towards our average rating[9], even if funds with higher initial ratings in general retained a slight advantage at the end of three years. For example, for Morningstar data, funds rated 4-stars or 5-stars in September 2001 obtain on average a 3.1 rating in September 2004 versus 2.7 for those with initial ratings of 1-star or 2-stars.

.

[8] *See Kinnel (2005, 2006).*
[9] *The average rating is less than 3-star for S&P because half the funds were rated 1-star or 2-star.*

Chart 2

Successive average ratings based on the initial rating in Sept. 2001

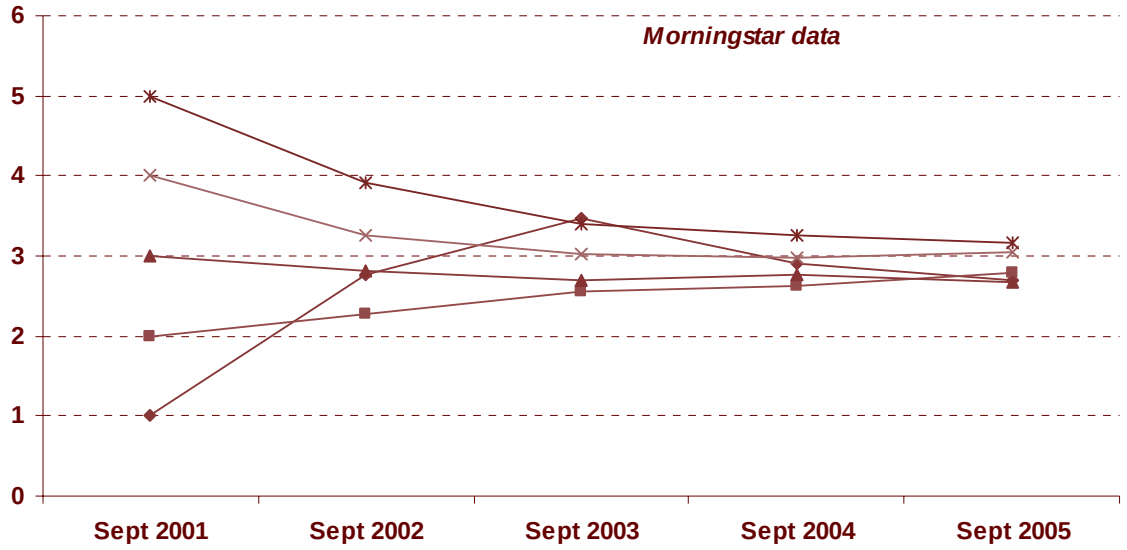
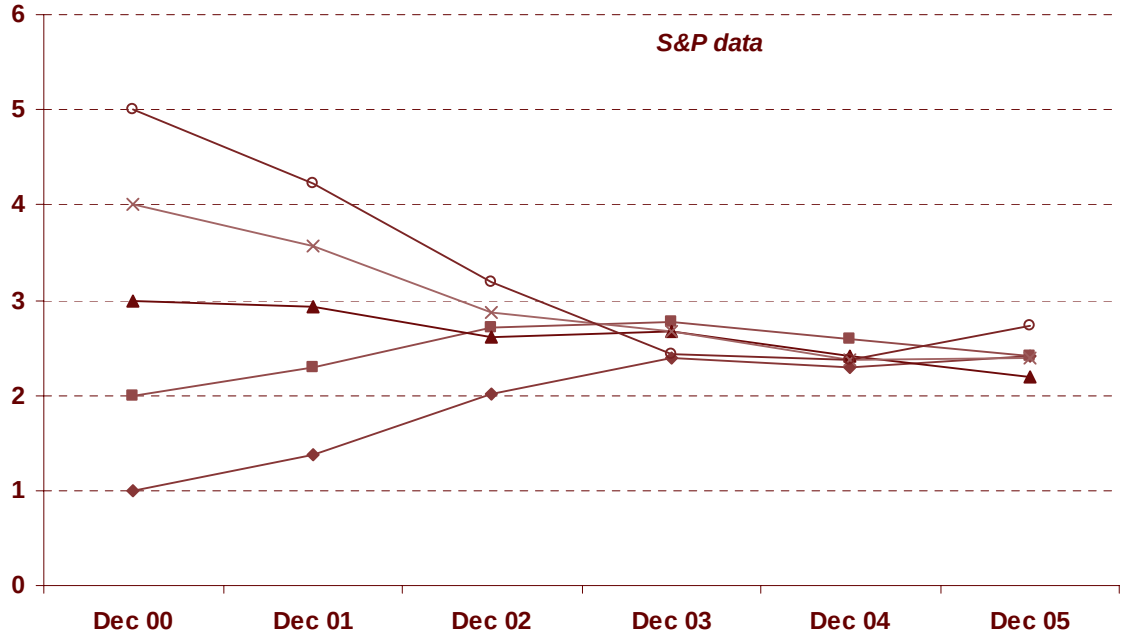


Chart 3

Successive average ratings based on the initial rating in Dec. 2000



Nevertheless, it is difficult to draw clear conclusions from these types of results:

- We have too few multiple-year periods to obtain sufficiently precise estimates and do not have the possibility to test the statistical significance of variances between average ratings of different groups;
- The above results are probably affected by a survivorship bias as funds that cease to be rated over the following years are not taken into account.

Table 1 below[10] illustrates the lack of clear conclusions, presenting the breakdown of funds according to their ratings over the previous three-year period, and introducing a category for unrated funds (NS).

Table 1

Standard & Poor's		**Rated in Dec. 2003**				
		1+2*	**3***	**4+5***	**NS**	**Total**
Rated in Dec. 2000	**1+2***	**38%**	**14%**	**23%**	**25%**	**100%**
	3*	**38%**	**21%**	**21%**	**19%**	**100%**
	4+5*	**46%**	**15%**	**22%**	**17%**	**100%**
	Total	**41%**	**16%**	**22%**	**21%**	**100%**

Standard & Poor's		**Rated in Dec. 2005**				
		1+2*	**3***	**4+5***	**NS**	**Total**
Rated in Dec. 2002	**1+2***	**46%**	**12%**	**6%**	**36%**	**100%**
	3*	**46%**	**20%**	**5%**	**29%**	**100%**
	4+5*	**32%**	**12%**	**36%**	**21%**	**100%**
	Total	**42%**	**13%**	**15%**	**30%**	**100%**

[10] *While this table is presented here only in the case of S&P data, the results would be qualitatively comparable in the case of Morningstar data.*

This table illustrates the following points:

- The percentages obtained vary significantly according to their starting date. Funds initially rated 4-stars or 5-stars in December 2002 have a significantly higher probability of being better rated three years later. This is however not the case if one takes December 2000 as the starting date (corresponding to the burst of the stock market bubble that may explain in part why many of the top performers of 1998 to 2000 failed to retain their rating over the following three years). For this reason, it would be useful to test the temporal stability of the results;
- The percentage of unrated funds three years later is relatively significant (approximately 20% to 30%), and is moreover higher among those funds with initially less favorable ratings. This observation highlights the need to consider the survivorship bias when studying the persistence of ratings.

4-2 Estimates based on monthly transition matrices

To surmount these challenges, we adopted the stochastic method that involves estimating monthly transition matrices:

- Recourse to high frequency information (i.e., monthly) improves the statistical accuracy of estimates for probabilities of changing from one star rating to another;
- The introduction of a category for unrated funds in such a matrix makes it possible to take into account funds included or excluded from the selection of rated funds, thus offsetting a potential bias.

4-2-1 The estimation method

To produce the transition matrices, the preliminary phase consisted in defining the different possible states of a fund. We retained six including the five star-rating levels and one for unrated funds (no stars or “NS”). Every fund of the dataset can consequently be rated monthly according to these six classifications.

The subsequent steps consisted in calculating for each month the number of changes from one rating level to another and then estimating the probabilities of transition from one rating level to another. For the following:

- n_{ij}^{t} the number of funds changing from level i to level j over the period t
- p_{ij}^{t} the probability of transition from level i to level j at period t

where i,j $\in$ {*, **, ***, ****, *****, NS}.

The transition matrix P^t for month t (that includes 6 lines and 6 columns) is made up of probabilities p_{ij}^t and is consequently represented as follows :

$$P^t = \begin{vmatrix} & NS & * & ** & *** & **** & ***** \\ NS & & & & & & \\ * & & & & & & \\ ** & & & & p_{ij}^t & & \\ *** & & & & & & \\ **** & & & & & & \\ ***** & & & & & & \end{vmatrix}$$

The estimation of the probability of transition is made on the basis of the likelihood of the dataset that equals on date t:

$$L_t = \sum_i \sum_j n_{ij}^t \; Log\left(p_{ij}^t\right)$$

We then calculate the estimator of maximum likelihood of probabilities of transition subject to habitual constraints:

$$\begin{aligned} &\hat{p}_{ij}^t \geq 0 \\ &\sum_j \hat{p}_{ij}^t = 1 \end{aligned} \qquad (1)$$

This estimator is represented by the following formula:

$$\hat{p}_{ij}^t = \frac{n_{ij}^t}{\sum_j n_{ij}^t} \qquad (2)$$

and is asymptotically unbiased with normal distribution. This standard variation is represented as follows:

$$\hat{\sigma}_{p_{ij}^t} = \left[\frac{\hat{p}_{ij}^t \left(1 - \hat{p}_{ij}^t\right)}{\sum_j n_{ij}^t} \right]^{1/2} \qquad (3)$$

This is a conventional estimator frequently used for example in connection with credit rating of issuers. However, fund ratings adhere to a particular logic that makes this estimator unsuitable. Indeed, while credit ratings have an absolute nature (the breakdown of issuers between different ratings is not predefined), the rating of funds is carried out on a purely relative basis: by construction, the proportion of funds belonging to each category is fixed. However, the estimator produced by (2) in no way imposes such a constraint.

It should however be noted that this problem applies only to the dataset of S&P and not Morningstar. Indeed, in the Morningstar rating mechanism, in contrast to S&P, ratings are calculated not only on the basis of funds distributed in France, but rather those distributed throughout Europe. Consequently, the proportion of funds distributed in France at each Morningstar star rating level is not predetermined.

For the S&P dataset, it is necessary to modify the preceding estimator as follows. At a given moment t, if one notes λ the proportion of rated funds in a database and (1-λ) the proportion of those not included, the proportion of funds in the different rating categories is consequently determined by the following vector (given that, in the S&P rating mechanism, 25% of rated funds receive 1 star, 25% two stars, 20% three stars, etc.):

$$X(\lambda) = \begin{bmatrix} 0.25\lambda & 0.25\lambda & 0.2\lambda & 0.2\lambda & 0.1\lambda & 1-\lambda \end{bmatrix}$$

By construction, at the following period t+1, there must exist a parameter λ' that defines a vector X(λ') of a similar structure, so that the proportion of rated funds remains unchanged.

The fact that the rating of funds is relative implies the existence of a pair (λ ,λ') such as :

$$X'(\lambda') = X'(\lambda) P^t \qquad (4)$$

There are accordingly five additional linear constraints for coefficients, in addition to the usual conditions needed to construct a transition matrix. This gives the estimators $\tilde{p}_{ij}^{t}$ by maximizing the likelihood L_t, under this entire set of constraints.

4-2-2 Test of the temporal stability

At this stage, it is possible to estimate the monthly transition matrices. Questions concerning the temporal stability of these matrices may be raised for two reasons. Firstly, if it turns out that these matrices were unstable (because, for example, of the high dependence of probabilities of monthly transition vis-à-vis the stock market cycle), it would be impossible to extrapolate the resulting estimations over several years. Secondly, the stability of these matrices affects the accuracy of estimations of probabilities. Indeed, when one works on monthly transitions, the probability of a fund remaining in the same rating category is very high (p_{ii} ~ 1) whereas the probabilities p_{ij} (with i≠j) that it migrates to another rating category are low. Thus if there are only a few funds in a different rating category, the evaluation of probability of migrating to another rating category is highly imprecise. For example if p_{ij}= 0.1 with N=25 observations, we obtain on the basis of (3) a standard deviation $\hat{\sigma}_{p_{ij}^t} = 0.06,$ that is to say same range of transition probability, which cannot be therefore considered significantly different from zero. This highlights the benefits of having stable transition matrices over time. Moreover, this allows for the use of more observations (N increases in our example and therefore the standard deviation decreases), resulting in more accurate estimates.

Temporal stability is traditionally tested by comparing the likelihood obtained under the assumption that matrices $\tilde{P}^t$ are:

(i) variable over time (with the log-likelihood thus called LV_{nc});

(ii) constant: $\tilde{P}^t = \tilde{P}\ , t = 1,...,T.$ (with the log-likelihood thus called LV_c).

.

This test must be slightly modified when constraint (4) is imposed. Under this constraint, each matrix $\tilde{P}^t$ contains (6-1)(6-1) = 25 independent coefficients per period. This number increases to 6(6-1) = 30 parameters without this constraint.

For the S&P dataset, where this constraint must be applied (4), the number of months of observations is T = 60 (December 2000 to December 2005). This results in an estimate of 1,500 parameters (60x25) if the matrices are distinct, compared to 25 parameters otherwise. The assumption of stability is thus rejected at a given threshold α = 0.05 if:

$$D = \left| LV_c - LV_{nc} \right| \geq \chi^2_{0.05}(1475) = 1565$$ [11]

[11] *In keeping with common practice, to calculate thresholds, the law $\chi^2(u)$ has been approximated by normal law $N(u, (2u)^{1/2})$.*

With our S&P dataset, this deviation is D =1038 and the assumption of stability for transition matrices over December 2000 – December 2005 is thus accepted.

For the Morningstar dataset, we did not apply constraint (4) and the number of months of observations is T = 47 (October 2001 to September 2005)[12]. This results in an estimate of 1410 parameters (47x30) if the matrices are distinct versus 30 parameters in the contrary scenario. The assumptions of stability would thereupon be rejected at a given threshold α = 0.05, if:

$$D' = \left| LV_c - LV_{nc} \right| \geq \chi^2_{0.05}(1380) = 1466$$

With our Morningstar dataset, this deviation is D'=1203 and the assumption of stability for transition matrices over October 2001 – September 2005 is thus accepted.

It should be noted that this stability of transition matrices offers a significant practical advantage. For an initial breakdown X_0 of funds, the final breakdown X_t, for a given time span t can be easily calculated through the following formula:

$$X_t' = X_0' \tilde{P}^{(t)}$$

where $\tilde{P}^{(t)}$ here represents matrix $\tilde{P}$ to the power of t.

In this way, it is possible to analyze the matrices for a given period of time, even if they exceed that of the database used. It should however be emphasized that this practice implies not only stability for the transition matrices over the period under review (which has been tested) but also beyond (which of course has not). In addition, as addressed in the Appendix 2, the situation of the matrices at the time t must represent all information necessary to evaluate the transitions. This technical discussion is presented in Appendix 2. Concretely, it could result in overestimating the stability of the transition matrices beyond one month.

[12] *When a test is conducted over a longer period (March 2001-March 2006), the assumption for stability is not verified for reasons given in footnote 3.*

4-2-3 Result of estimations

Table 2: One month transition matrix estimated on the basis of the S&P dataset

Initial\Final	NS	*	**	***	****	*****
NS	97.14%	0.41%	0.62%	0.54%	0.77%	0.52%
	0.16%	*0.06%*	*0.08%*	*0.07%*	*0.09%*	*0.07%*
*	1.26%	84.84%	13.48%	0.40%	0.02%	0.00%
	0.16%	*0.51%*	*0.48%*	*0.09%*	*0.02%*	*0.00%*
* *	0.78%	14.23%	67.3%	16.96%	0.73%	0.00%
	0.12%	*0.48%*	*0.72%*	*0.51%*	*0.12%*	*0.00%*
* * *	0.70%	0.84%	22.34%	62.30%	13.75%	0.07%
	0.13%	*0.14%*	*0.63%*	*0.84%*	*0.52%*	*0.04%*
* * * *	0.66%	0.00	0.88%	15.23%	76.68%	6.55%
	0.12%	*0.00%*	*0.14%*	*0.55%*	*0.70%*	*0.38%*
* * * * *	0.38%	0.00%	0.00%	0.05%	14.70%	84.87%
	0.13%	*0.00%*	*0.00%*	*0.05%*	*0.78%*	*0.79%*

Numbers in italics correspond to estimators of standard deviations

Table 3: One month transition matrix estimated on the basis of the Morningstar dataset

Initial\Final	NS	*	* *	* * *	* * * *	* * * * *
NS	96.47%	0.25%	0.68%	1.48%	0.72%	0.40%
	0.17%	*0.05%*	*0.08%*	*0.12%*	*0.08%*	*0.06%*
*	0.26%	86.89%	11.82%	0.78%	0.19%	0.06%
	0.13%	*0.87%*	*0.82%*	*0.22%*	*0.11%*	*0.06%*
* *	0.40%	3.44%	84.47%	11.36%	0.31%	0.02%
	0.08%	*0.24%*	*0.49%*	*0.42%*	*0.07%*	*0.02%*
* * *	0.19%	0.06%	6.75%	88.79%	4.18%	0.03%
	0.04%	*0.02%*	*0.24%*	*0.31%*	*0.19%*	*0.02%*
* * * *	0.23%	0.00%	0.20%	10.42%	85.50%	3.64%
	0.06%	*0.00%*	*0.06%*	*0.41%*	*0.49%*	*0.25%*
* * * * *	0.37%	0.05%	0.05%	0.21%	13.78%	85.53%
	0.14%	*0.05%*	*0.05%*	*0.10%*	*0.79%*	*0.81%*

Numbers in italics correspond to estimators of standard deviations

Estimates for monthly transition matrices are presented above in tables 2 and 3. These tables highlight the high degree of monthly mobility among different categories:

- In the case of S&P data, 15% of 5-star rated funds lose this rating each month and one third of 2-star or 3-star funds change their rating (see table 2);
- In the case of Morningstar data, approximately 15 % of funds change their star-rating (see table 3).

Stars are calculated on the basis of performances recorded over the preceding 36 rolling month period. One would have expected in consequence a greater degree of stability in ratings from one month to another. To explain this somewhat surprising result on an initial reading, a more in-depth analysis is necessary.

5 - Analysis and interpretation of results

Three key questions should be addressed:

- In light of the high monthly mobility for star ratings, on the basis of what time span can we consider that the initial fund rating no longer has a significant impact on the final rating?
- Do past ratings offer additional information in relation to information already provided by current ratings?
- What are the implications of the addition and removal of funds in the dataset of rated funds?

5-1 The level of rating persistence

5-1-2 What is the level of persistence at a 3 year interval?

Rating categories should be compared at three-year intervals as this is the time span for which historical data used to calculate star ratings no longer overlaps.

As already indicated, the three-year transition matrix can be determined by multiplying the monthly transition matrix by 36. The results are presented in tables 4 (S&P) and 5 (Morningstar). In both cases, funds with an initial 5-star rating appear to have a higher probability than lower-rated funds to be rated 4-stars or 5-stars three years later. However, in absolute terms, there is nothing exceptional about this probability. For example, for S&P data it is slightly less than 28% (or 32% if funds no longer rated are eliminated) resulting in the same scale order of magnitude as the percentage of funds that S&P ranks each month in the 4-star to 5-star category (30%). In other words, initial 5-star funds do not perform much differently than other funds from the same time period at the end of three years (their breakdown between the different categories of star ratings is relatively similar to that retained by S&P for the entire population of rated funds). They differ only from funds initially rated less favorably because of the persistence of the lower ratings of the latter.

Table 4: Three-year transition matrix estimated on the basis of the S&P dataset

Initial\Final	NS	*	* *	* * *	* * * *	* * * * *
NS	42.2%	13.3%	13.9%	11.6%	12.3%	6.6%
*	18.0%	22.6%	21.5%	16.2%	14.9%	6.7%
* *	17.1%	22.3%	21.5%	16.5%	15.5%	7.2%
* * *	16.5%	21.8%	21.3%	16.7%	16.1%	7.7%
* * * *	15.6%	20.9%	21.0%	16.9%	17.0%	8.5%
* * * * *	14.4%	19.8%	20.6%	17.3%	18.2%	9.6%

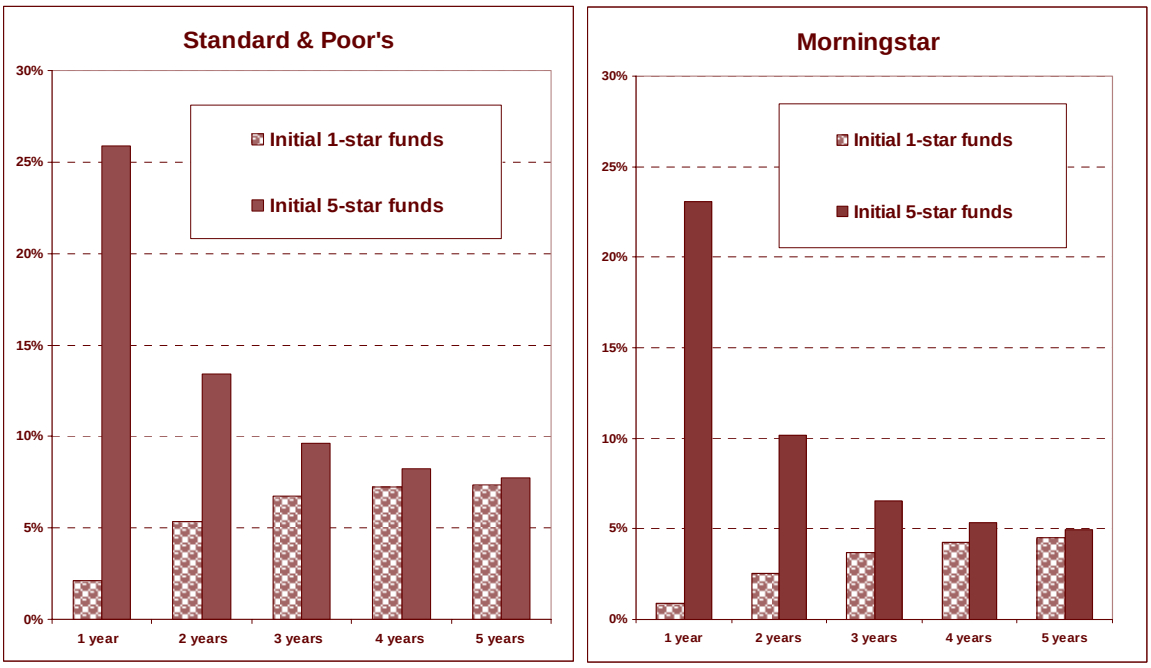
Table 5: Three-year transition matrix estimated on the basis of the Morningstar dataset

Initial\Final	NS	*	**	***	****	*****
NS	30.1%	4.6%	16.7%	30.5%	13.8%	4.3%
*	5.5%	8.2%	25.8%	41.5%	15.4%	3.7%
* *	5.4%	7.2%	24.7%	42.0%	16.5%	4.2%
* * *	5.1%	6.6%	24.0%	42.2%	17.4%	4.6%
* * * *	5.1%	5.8%	22.5%	41.9%	19.1%	5.6%
* * * * *	5.3%	5.1%	21.0%	41.3%	20.8%	6.6%

5-1-2 According to what time span does the initial rating cease to have an impact?

The initial responses to this question can be obtained by first comparing the probability for a fund to be rated 5-stars for a given period of time, depending on whether it was initially rated 5-stars or 1-star. As shown below in chart 4, an initial 5-star fund has significantly greater chances than an initial 1-star fund to be rated 5-stars within a time span of less than 3 years. However, this advantage quickly decreases, tending to disappear after 4 or 5 years.

Chart 4 : Probability for initial 5-star rating of a UCITS



N.B.: Results obtained for S&P and Morningstar are not directly comparable primarily because breakdowns for stars and estimation periods are not identical.

For a more precise idea of the time span for which the initial rating no longer is meaningful, the uncertainties concerning estimates for the probability of transition must be taken into account. This can be done by means of the following method.

Consider the transition matrix $\tilde{P}^{(t)}$ with a given time span t. Thus, $\tilde{P}_{i,j}^{(t)}$ the probability of a fund with an initial rating in category i two and in category j at time t. The standard deviation $\tilde{\sigma}_{\tilde{p}_{ij}^t}$ of these probabilities depends in a complex fashion on all standard deviations $\hat{\sigma}_{p_{ij}^t}$ of matrix $\tilde{P}$. To evaluate them, we have used the Monte Carlo method. After completing this calculation, we consider that the initial rating i or i' of a fund has no impact, at threshold α, for the probability of being rated j, when:

$$\left| \frac{\tilde{p}_{i,j}^{(t)} - \tilde{p}_{i',j}^{(t)}}{\tilde{\sigma}_{\tilde{p}_{ij}^t - \tilde{p}_{i'j}^t}} \right| < c_\alpha \quad \textit{for every } j$$

where c_α corresponds to the Student test at threshold α.

The results are summarized below in table 6. For example, for Morningstar, a fund with an initial 5-star rating at the end of a 46 month period becomes a UCITS with an initial 1-star rating, while after only 26 months, a fund with an initial 4-star rating, i.e. before the same 3-year period required to obtain a new historical performance completely independent of the data used for the initial rating.

Table 6: Time span within which a UCITS with an initial 5-star rating will become statistically equivalent (at a 5% threshold) to a fund with an initial rating of…

	S&P	Morningstar
4*	26 months	26 months
3*	35 months	39 months
2*	39 months	43 months
1*	42 months	46 months

N.B.: Results obtained for S&P and Morningstar are not directly comparable primarily because breakdowns for stars and estimation periods are not identical.

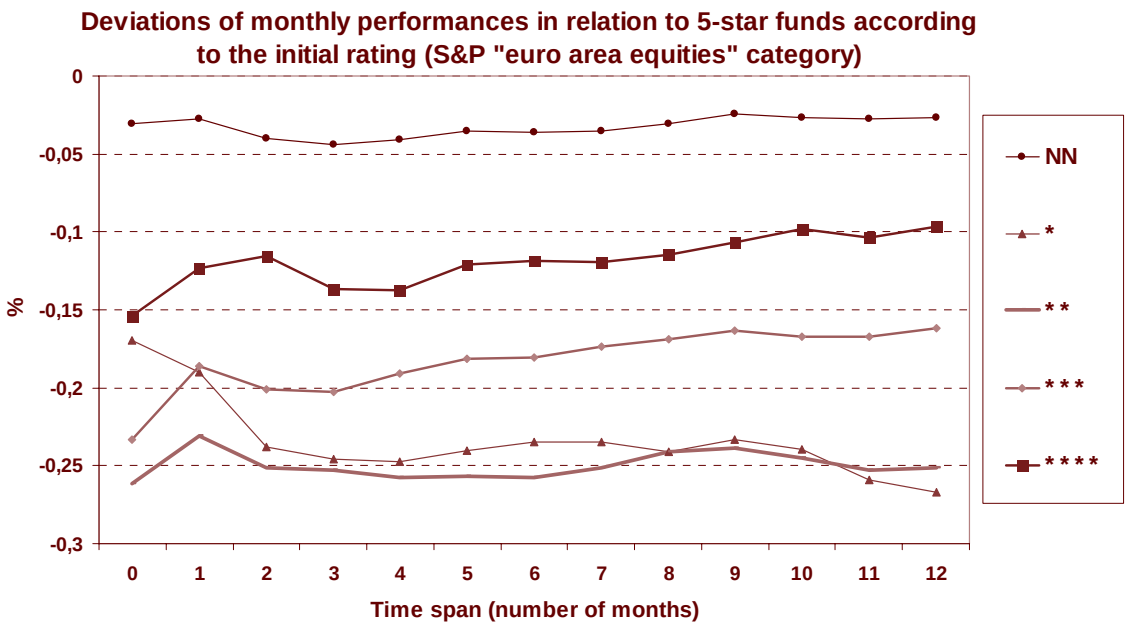
For this critical three-year time span, it is nevertheless reassuring to note that funds with an initial 5-star rating retain an advantage over funds with initial ratings of 1-star or 2-stars. However, this advantage is not maintained up to 4 years.

5-1-3 What are the variances in performance?

Up to this point, our analysis deliberately limited its focus to the persistence of ratings. However, to establish the orders of magnitude, a comparison of future returns of funds (non-risk-adjusted) according to their initial rating could be useful.

For the S&P dataset limited exclusively to "euro area equity" funds, chart 5 presents below variances of monthly performances in relation to 5-star funds according to the initial rating alone[13]. This variance appears relatively stable over a 12 month period. For example, funds with an initial 5-star rating tend on average to outperform those with initial ratings of 1-star or 2-stars by approximately 25 basis points per month over the 12 months following this rating. However, given the scope of standard deviations, this differential in returns is not statistically meaningful, even though it is not negligible in absolute value terms. The performance of a randomly selected 5-star fund is not significantly higher than that of any randomly selected fund with a lower-rating. In other words, an investor wishing to take advantage of this return variance can only do so if he is a "buyer" of all 5-star funds and a "seller" of all the lower-rated funds (bearing in mind that all associated transaction costs must be taken into account).

Chart 5



[13] *A calculation is accordingly made, for each fund, at period t=0 in a given category, its return at t+1, t+2, …, t+12 months. The average return is then inferred for all funds of a given category at time t for these different time spans independently of their subsequent ratings. For each category, the average of this monthly indicator between 2001 and 2004 is then calculated.*

5-2 The impact of historical star rating data

Estimations of the preceding transition matrices involve an implicit assumption that past ratings played no role and in the same manner that the current rating summarizes all available information for the funds. Such an assumption is difficult to validate in this study because of the need for long-term data[14].

However, an attempt to evaluate the limits of such an assumption is possible. This requires proceeding on the basis of aggregated transition matrices, otherwise the number of transitions to a given rating is too low to be correctly estimated. We consequently reduced the number of categories to four (instead of six previously) by aggregating 1-star and 2-star ratings, on the one hand, and 4-star and 5-star ratings on the other. On the basis of this new classification, we have taken into account historical data of funds by calculating the probabilities of transition not only according to their current rating but also their rating 12 months before. Table 7 presents below this new transition matrix (on the basis of S&P data):

Table 7: One-month transition matrix according to current and past star ratings (-12 months)

Previous ratings (-12 months)	Current ratings	Future ratings (+1 month)			
		NS	1* ou 2*	3*	4* ou 5*
NS	1* or 2*	0.5%	89.9%	9.4%	0.3%
1* or 2*	1* or 2*	1.3%	92.5%	5.9%	0.2%
3*	1* or 2*	0.8%	83.0%	15.8%	0.4%
4* or 5*	1* or 2*	1.2%	77.9%	19.8%	1.2%
NS	3*	0.4%	23.3%	59.8%	16.5%
1* or 2*	3*	0.9%	29.0%	57.7%	12.3%
3*	3*	0.8%	25.1%	64.8%	9.3%
4* or 5*	3*	0.5%	16.6%	68.5%	14.3%
NS	4* or 5*	0.9%	0.4%	6.9%	91.7%
1* or 2*	4* or 5*	1.1%	1.1%	21.5%	76.2%
3*	4* or 5*	0.5%	0.3%	18.0%	81.1%
4* or 5*	4* or 5*	0.5%	0.5%	7.5%	91.5%

Source : Standard & Poor's dataset

This analysis highlights two significant differences according to past fund ratings. Firstly, a fund less favorably rated today, but rated higher in the past, has a better chance of rebounding in the future than consistently low rated funds. Similarly, a fund recently rated 4-star or 5-star, though previously rated 1-star or 2-stars, has a greater chance of being subsequently downgraded than an older star-rated fund.

These results appear to point to a force of rebound that tends to return funds to their previous rating when this has recently changed. This suggests that the historical data for star ratings of a fund offer useful information in addition to that provided by its current rating.

[14] *These are standard test in the econometric literature on the order of Markov chains.*

Table 7 also highlights another result. Among 4-star or 5-star funds, persistence is higher not only among those with such initial high ratings but also those not yet rated 12 months ago. The level of persistence is consequently characterized by a U curve with "seasoned funds" and "young funds" at the two ends. Between these two extremes, "rising stars" are in the most precarious position. It is better to be dropped directly down from the top than promoted up from the bottom …

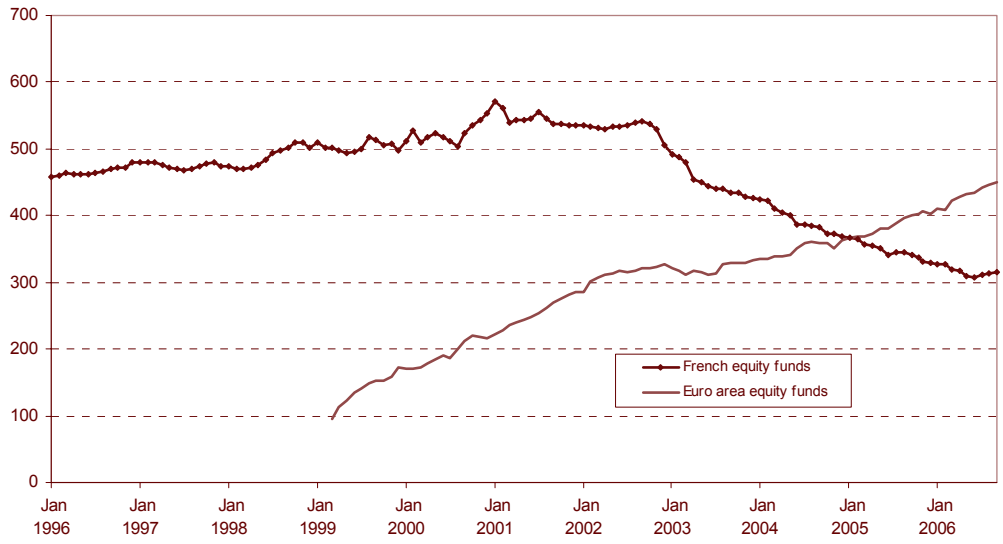
5-3 Impact of entering and exiting fund ratings

5-3-1 The scope of entries and exits

A significant phenomenon is the relatively high number of funds which, from one month to the next, become 'starred' (entries) or, to the contrary, that cease to be so (exits). Thus for S&P, monthly entries represent approximately 1.3% and exits 0.8% of existing funds. Overall, over one year 8.1% have disappeared from ratings and 13% were recently rated. For Morningstar, figures are comparable for entries (13.6% annually) but significantly lower for exits (2.7% annually)[15].

The high number of entries partially reflects strong growth in the number of newly created equity funds in France. The introduction of the euro combined with the strong bull market for equities in the late 90s contributed to the growth of new "euro area" funds as chart 6 shows below. A three-year lag resulted mechanically in significant entries into the S&P and Morningstar fund ratings in the first half of the current decade.

Chart 6
Number of UCITS distributed in France



Source : AMF

[15] We note however that these results for the Morningstar dataset are valid only between October 2001 and September 2005. Especially as indicated in footnote 5, entries into rated funds increased significantly during the period due to the introduction of new fund categories by Morningstar.

However the magnitude of the flows of entries and exits also reflects the migration of funds to different categories for a number of possible reasons (change in fund benchmark, migration to a new category better adapted to the fund's characteristics, etc.). In particular, following the changeover to the euro, a significant number of "French" equity funds migrated to the "euro area" category. This trend could also help explain the lower rate of outflows for the Morningstar universe versus S&P as Morningstar does not distinguish between these two categories.

5-3-2 Is there an incubation bias?

As shown below in table 8, newly rated funds appear to have above-average ratings with a statistically meaningful difference both for S&P and Morningstar. For example, with S&P, close to 18% of recently rated funds are granted a 5-star rating, compared with 10% for the entire population. Similarly only 14% of new funds are rated 1-star versus 25% for the entire population.

This may be subject to different interpretations. However, the hypothesis of an "incubation bias" potentially associated with the strategic behavior of management companies cannot be excluded. Indeed, these companies may be tempted to improve the standing of their newly created funds by closing or changing the category of the lower performers before they attain the key milestone of three-year historical data. Pressure of demand could contribute to these types of adjustments if new lower performing funds fail to attract sufficient assets. In addition, it cannot be entirely excluded that some funds fail to be rated while their performances appear to be mediocre because the relevant data is not transmitted to the rating agencies.

Tableau 8

Rating for new stars (S&P)

Rank	1*	2*	3*	4*	5*
Proportion	25%	25%	20%	20%	10%
New stars	14.10% (1)	22.10%	19.20%	26.80%(1)	17.80%(2)

Rating for new stars (Morningstar)

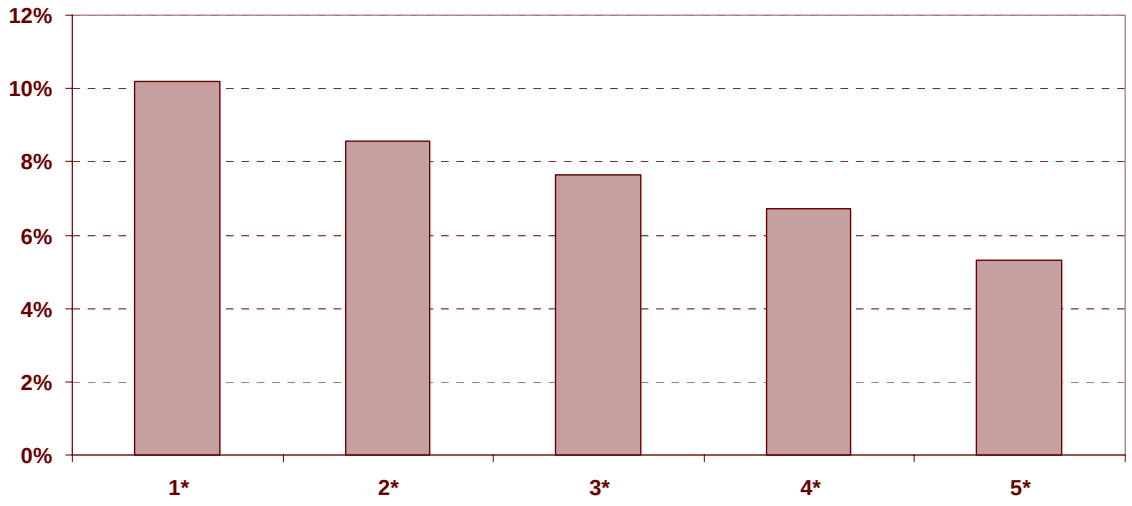
Rank	1*	2*	3*	4*	5*
Average	6%	22.40%	42.70%	21.50%	7.40%
New stars	7%	19.20%	42.10%	20.30%	11.40%(2)

(1) Meaningful at 5% threshold *(2) Meaningful at 1% threshold*

5-3-3 Is there a survivorship bias ?

Chart 7

Probability of outflows based on the number of star-ratings obtained in the 12 prior months



Source : Standard & Poor's dataset

While the existing literature includes little analysis concerning the existence of an incubation bias, the role of the survivorship bias on fund performance is in contrast well-known and documented. For example, studies of the US mutual fund market clearly demonstrate that low-performing funds have a reduced probability of surviving.

This tendency has been in part confirmed in our study, though in a somewhat unexpectedly the survivorship bias appears less pronounced than the incubation bias.

Indeed, in the S&P database, exits concern in priority the lowest-rated funds (cf. chart 7). For a one-year period, the probability of a losing a star-rating is twice as high for a fund with an initial 1-star rating than for a fund with an initial 5-star rating (slightly more than 10% versus above 5%). However:

- This result is not found with the Morningstar data where the probability of becoming unrated does not depend significantly on the initial number of stars.
- By examining the S&P data in more detail, one notes that funds lose star-ratings not because of their disappearance but rather their migration to other categories. It would be more accurate in consequence to speak of a "migration bias" rather than a survivorship bias. As already indicated, this is in part explained by the changeover to the euro when a number of funds changed category (from

"French" to "euro area") resulting for many in the loss of their S&P rating[16]. This migration bias would be in this case temporary and specific to the S&P rating mechanism (resulting from the distinction between "French" and "euro area" funds in its classification system). However, there are other types of migrations to categories such as "SRI/Ethical"[17] or "balanced funds" that are more difficult to interpret. These changes may be related not only to changes in fund benchmarks but also reclassifications by S&P if these funds were initially rated in a category considered less pertinent.

In any case, the selection bias appears more pronounced when star-ratings are received rather than lost. This is of some importance because these two types of bias do not operate in the same direction. The selection bias for funds receiving a star-rating pushes down the ratings of funds already rated, while the opposite occurs upon exit. The prevalence of the incubation bias may consequently partially explain one of the results identified in our study:

- A fund with an initial 5-star rating does not have, after three years, significantly greater chances than a randomly selected fund (including those unrated three years before) to be rated 4-stars or 5-stars;
- Nevertheless, a fund with an initial 5-star rating has greater chances of remaining well rated after three years than a fund with an initial low-rating.

[16] When a fund migrates from the category of "French equities" to "euro area equities" it retains a star-rating only if S&P accepts the historical data on performances when it changes category, which is not automatic.

[17] The "Ethical/SRI" category was introduced by S&P in October 2003. The funds migrating to this category retained their star-rating only in those cases where S&P considered that they were already SRI managed during the three preceding years.

6 - Conclusion

Several relatively robust results are provided by the above analysis concerning the level of persistence for star ratings assigned to equity funds (French or euro area) distributed in France.

Firstly, there is certain level of persistence for Morningstar or S&P star ratings, though it is low and above all tends to rapidly disappear beyond a three-year period (whereas, according to studies conducted by the Banque de France, investments in UCITS are held on average slightly more than four years). Consequently, a rebound force over the long term pushes leading funds back up to the top ratings initially held. Indeed, variances of future performances between initially high-rated and low-rated funds are on average significant, though their high level makes it in practice difficult for investors to take advantage of them (except for "buyers" of all 4-star or 5-star funds and "sellers" of all 1-star or 2-star funds). It is nevertheless possible to identify cases where the level of persistence is particularly high and can be exploited by investors. On the one hand, funds consistently highly-rated in the past (i.e. "seasoned stars") have a greater of probability of retaining their rating level (in relation to those recently upgraded from the lower-rated funds). On the other hand, funds with initial high ratings (that we refer to as "young stars"), i.e. those recently rated (after passing the historical three-year milestone) and included directly in the top-rated category had the greatest chances of retaining their position. In other words, to use an analogy of the ATP ranking for tennis players, one has better chances by betting on Roger Federer, who has been the top-ranked player for several years, than on a player who has suddenly reached the top 10 after winning one or two tournaments, except if the latter is a young player who only recently joined the professional tennis circuit.

The second observation concerns the importance of entries and exits among rated funds and the impact of these changes on the latter. When ratings are based on performances for the preceding 36 rolling months, it may be inferred that the ratings are highly stable from one month to another or even one year to the next. However, this has not been demonstrated. The entry of a significant number of new funds (at an annual rate averaging 13% over the period reviewed) has a major impact on ratings, especially as the new arrivals are generally better rated, which automatically downgrades the older funds. In contrast, even though there is an overweighting of low-rated funds among funds that cease to be rated, the impact of the exit selection bias on ratings appears to be less than that of entries. This observation is noteworthy, as previous studies that focus primarily on US and UK markets have analyzed above all the survivor bias, neglecting what we have referred to as an "incubation bias".

These results must however be interpreted with caution. In the first place, as previously noted, star-ratings are not designed to serve as the sole criteria for selecting funds. Their limited persistence in consequence only confirms the evident need to consider many other quantitative and qualitative criteria. Furthermore and above all, our results are based on short-run historical data (2000-2005) and the instability of star ratings could be accentuated by factors specific to this period. On the one hand, there was considerable upheaval in financial markets, including very pronounced reversals in market sectors and prevailing investment styles.

On the other hand, the French market for equity funds began to take form and develop only in the late 1990s (as demonstrated by their high rate of entry into fund ratings). This market had consequently not yet attained maturity in the period covered by this study. Even if the data on fund ratings already offers a rich source of information, it will only gain pertinence if it is expanded to cover a longer period and the French market for equity funds becomes more mature.

From our results, two simple recommendations can be made to improve the robustness and usefulness of the information on fund ratings for investors:

- **Increase transparency about entries and exits into fund ratings**: a systematic inventory of incoming funds and better knowledge of the reasons that funds are exited would improve information quality;
- **Provide historical data for previously rated funds** and not only stars from the last available month since it has been demonstrated that a fund consistently rated in the past has greater chances of remaining so in the future[18] (excluding the special case of new arrivals with initial high ratings).

[18] This does not lead us to recommended star-ratings based on historical performance data for periods exceeding three years, as is done by Morningstar in the United States (where star-ratings are granted according to weighted averages of ratings over 3, 5 and 10 years for funds with available historical data for longer periods). As demonstrated by Morey (2002), it introduces an age bias in the comparison of funds: on the basis of identical underperformance over recent years, a fund over 10 years old is less likely to lose its star rating than a fund less 5 years old.

References

Amenc N., Le Sourd V. (2005), "*Rating the Rating*", Edhec Risk and Asset Management Research Center, Working Paper.

Blake C., Morey M. (2000), "*Morningstar Rating and Mutual Fund Performance*", Journal of Financial and Quantitative Analysis, 35, 3, 451-83.

Del Guercio D., Tkac P. (2001), "*Star Power: the Effect of Morningstar ratings on Mutual Fund Flows*", Federal Reserve Bank of Atlanta, Working Paper 2001-15.

De Marchi R. (2006), "*La persistance des performances des OPCVM actions françaises*", Banque et Marchés N°82.

Giles T., Wilsdon T., Worboys T. (2002), "*Performance persistence in UK equity funds – A literature review*", Charles River Associates.

Kinnel R. (2005), "*Rating the Star Rating*", Morningstar FundInvestor (December).

Kinnel R. (2006), "*The Star Rating Gains Momentum*", Morningstar FundInvestor (November).

Khorana A., Nelling E. (1998), "The *Determinants and Predictive Ability of Mutual Fund Ratings*", Journal of Investing (November), 61-66.

Morey M. (2003), "*The Kiss of Death: a 5-star Morningstar Mutual Fund Rating?*", Journal of Investment Management, 3,2, 41-52.

Morey M. (with Gottesman A.), (2006), "*Morningstar Mutual Fund Ratings Redux*", Lubin school of Business, Pace University, Working Paper.

Appendix 1

Fund ratings:
What is the impact of risk adjustments?

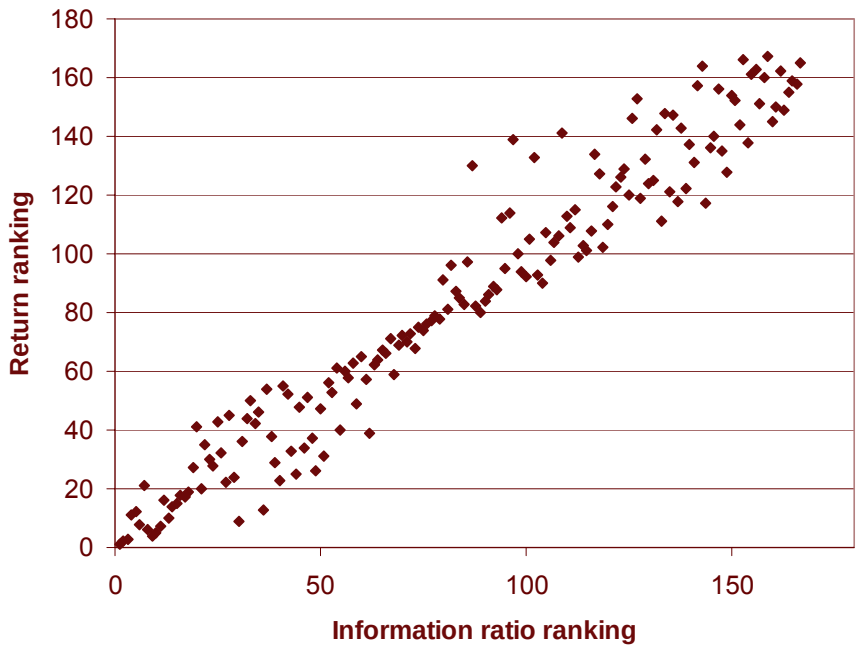
Ratings issued by S&P and Morningstar are based both on a comparison of returns, as well as, though according to different methods, risk-adjusted measures. This raises an inevitable question regarding the link between ratings based on risk-adjusted returns ratings and non-risk-adjusted returns, i.e. a direct comparison between the returns of funds over the preceding three year period.

This study may contribute to answer this question in the case of S&P data. In effect, for S&P-rated funds in December 2005, we possess not only their star-ratings but also:

- the value of the information ratio[19] used to calculate this rating,
- the non-risk-adjusted return for the previous three-year period.

The following chart that covers the population of "euro area" equity funds compares ratings according to these two criteria. On this basis, these two ratings appear to be very close.

Rankings according to returns and the information ratio



[19] *The indicator used by S&P is an information ratio that is not calculated on the basis of the benchmark index of the fund in question but rather for the average return for funds in the category.*

To quantify more precisely the correlation between the two ratings, we conducted a test of independence using the Spearman correlation[20]. This resulted in a rejection of the assumption of the 1% threshold of independence for both ratings.

The results of this study suggest that risk-adjustments have only a marginal impact on fund ratings. However, the limitations of this analysis should be emphasized:

- it concerns only a single period (ratings from the end of December 2005 calculated from performances over 2003, 2004 and 2005), and consequently not necessarily applicable to other periods;
- the same study was not carried out for the Morningstar ratings that uses a different type of risk-adjusted measure.

[20] The principle of this test is as follows. A comparison is carried out of ratings obtained using X and Y variables and the rankings R and Q, resectively. For each observation the statistic
$D_i = R_i - Q_i$ is calculated. The Spearman test is thus based on the following statistic:

$$\tau = 1 - \frac{6}{n\left(n^2 - 1\right)} \sum_{i=1}^{n} D_i^2$$

where n is equal to the number of observations

The critical region of the test of positive dependency between the two ratings is of the type $\tau > c$, with c varying in function of the threshold used.

Appendix 2

Specific bias created by transition matrices

Analysis of transition matrices offers the advantage provided by full use of monthly information on inflows and outflows of the universe of rated funds. However, because of insufficient data about their temporal dimension, certain biases specific to this method cannot be eliminated.

To understand these biases, consider a simple example where all funds have an expected return of zero because over a given period, this return may have an equal probability of being valued r or -r. It is inferred that returns are different funds are independent (both temporally and between each other) and the rating depends simply on the average return over the two periods (present and past).

According to these assumptions, funds could be classified in three categories, rated C_r, C_0 and C_{-r}, corresponding to average returns equal to respectively r, 0 and -r.

The transition matrix for the period would thus equal

$$P_1 = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{bmatrix}$$

Given the independence of the returns, the transition matrix for two periods equals

$$P_2 = \begin{bmatrix} 0.25 & 0.5 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

You can then be easily verified that $P_2 \neq P_1 * P_1$ since:

$$P_1 * P_1 = \begin{bmatrix} 0.375 & 0.5 & 0.125 \\ 0.25 & 0.5 & 0.25 \\ 0.125 & 0.5 & 0.375 \end{bmatrix}$$

The explanation for this result is as follows. The transition matrix P_1 is unsuited because it does not take into account certain information. More specifically, consider a fund rated C_0 at time t. If for the previous period it was rated C_r, its present return would be -r. For the following period, it would accordingly have the same probability ½ of being rated C_0 and C_{-r}. In contrast, if previously rated C_{-r}, its present return would be r and for the following period, it would have the same probability of being rated C_0 and C_r. In other words, as the rating takes into account two periods, a fund whose rating improves has a higher probability of being better

rated in the future. A suitable transition matrix must consequently take into account this information that in this case would give $P_2 = P_1 * P_1$. In this example, the transition matrix requires 4 states (instead of 3), distinguishing funds rated C_0 according to their previous return (r or -r).

This result could be widely applied without difficulty. Consequently, our study should have taken into account all historical rating data. However, for reasons relating to the data and size of the resulting transition matrices, in practice this is not possible.

However, this failure to consider historical rating data does not adversely affect the results of our study. In effect, in section V.2 on the impact of historical data for star-ratings, we demonstrated a mechanism contrary to that described above and the origin of the bias mentioned. A recently upgraded fund has a higher probability of being subsequently downgraded. Furthermore, our estimations of transition matrices have indicated a surprisingly high mobility for ratings from one month to another that also does not corroborate the importance of this bias.

Publication Director:
Hubert Reynier h.reynier@amf-france.org
33 (0) 1 53 45 6365

Chief Editor:
Fabrice Pansard f.pansard@amf-france.org
33 (0) 1 53 45 6357

Secretariat:
Muriel Visage m.visage@amf-france.org
33 (0) 1 53 45 6335

Note — *Working Papers* is coordinated by the Research Department of the Autorité des marchés financiers. It contains research carried out in-house of by outside researchers from the academic world, with whom Autorité des marchés financiers cooperates on a regular basis, notably through the Scientific Advisory Board.

Articles in *Working Papers* express the personal views of their authors, which are not necessarily the views of the Autorité des marchés financiers.