OPEN ACCESS
Journal of Financial Management, Markets and Institutions
(2023) 2350001 (30 pages)

DOI: 10.1142/S2282717X23500019

ABSOLUTE OR RELATIVE: THE DARK SIDE OF FUND RATING SYSTEMS

EMANUELE MARIA CARLUCCIO
University of Verona, Italy
emanuelemaria.carluccio@univr.it

PAOLO ANTONIO CUCURACHI
University of Salento, Italy
paolo.cucurachi@unisalento.it

UGO POMANTE
Tor Vergata University of Rome, Italy
ugo.pomante@uniroma2.it

Received 12 December 2022
Accepted 11 April 2023
Published 31 May 2023

Academic literature and market practitioners have always devoted great attention to the analysis of asset management products, with particular regard to fund classification and performance metrics. Less attention has been paid to rating methodologies and to the risk of attributing positive ratings to underperforming asset managers. The most widespread rating criterion is the ordinal one, which is based on the assumption that the best asset managers are those who have performed better than their competitors regardless of their ability to achieve a given threshold (i.e. a positive overperformance against the benchmark). Our study, after a description of the most common risk-adjusted performance measures, introduces the idea of attributing the rating on a cardinal basis, setting in advance a given threshold that should be achieved to receive a positive evaluation (i.e. a rating equal to or higher than 3 on a scale of 1–5). The empirical test conducted on a sample of funds (belonging to the main equity and bond asset classes) made it possible to quantify the effects of the cardinal approach on the attribution of the rating and on the probability of assigning a good rating to underperforming funds. Empirical analysis also highlighted how the cardinal method allows, on average, better performance than the ordinal one even in an out-of-sample framework. The differences between the two methodologies are particularly remarkable in efficient markets such as the North American equity market. The two rating assignment systems were also analyzed using contingency tables to test the ability to anticipate the default event (underperformance relative to the benchmark). The policy suggestion emerging from our study concerns the significant impact of the rating criterion

in reducing the risk of recommending funds that, despite a good rating, have failed to perform satisfactorily and are unlikely to do so in the future either.

Keywords: Mutual funds; performance evaluation; rating; persistence.

JEL Classifications: G11, G23

1. Introduction

The academic literature on risk-adjusted performance measures is boundless and has proposed, over the years, a number of different techniques and methodologies in order to evaluate the quality of asset management products. The mandate given to an asset manager falls within the agency theory since the investor (principal) is delegating to the fund manager (agent) the task to replicate (for passive managers) or to outperform (for active managers) the market index. According to the agency theory (Starks 1987) incentive commissions — and the underlying performance measure — can align the interests of principal and agent when regulation and market forces fail to achieve this goal. Since Sharpe's pioneering 1966 article proposing a *return-to-variability-ratio* based on the CAPM assumptions, there has been a proliferation of measures aimed at evaluating mutual funds that, while maintaining the same arithmetic structure of the Sharpe's Ratio, adopt different measures of excess return and risk.

The main measures of *risk-adjusted performance* (e.g. the Sharpe ratio or the information ratio) have a precise financial significance since they express the *basis points* that the fund manager was able to generate (in excess of the return of the *risk-free* asset or of a benchmark) for unit of risk taken; however, there are also other indicators (e.g. the Morningstar Risk-Adjusted Return) whose numerical value does not assume any specific financial meaning and which can, therefore, be interpreted only in terms of relative ranking with respect to funds belonging to the same *peer group*. It is worth mentioning that the comparison of funds via a *risk-adjusted performance* indicator must be carried out on the basis of a single time window (the Sharpe ratio computed on the three-year period, 2020–2022 certainly cannot be compared with the same measure computed on the three-year period, 2017–2019) and for a homogeneous sample of products (an equity fund cannot be compared with a bond fund).[a] The relative assessment is certainly the one prevailing for the purposes of attributing a rating by the main independent mutual funds rating companies, as well as the one favored by the banks called to create — picking from the

[a]We are aware that performance metrics can have different purposes since it can be used to judge the ability of a fund manager (e.g. for remuneration purposes) or the efficiency of a portfolio and that sometimes the goal is to evaluate investments belonging to different peer groups. In the paper, we dont't address the problem of selecting the right asset allocation which is based on the evalutation of the efficiency of different asset classes (stocks, bonds, cash, etc.), but following Sharpe's teachings we assume a hierarchical approach in which the strategic asset allocation precedes the selection for each asset class of the best mutual fund. In this case, it is necessary to compare funds belonging to same peer group since the decision about the weight of each asset class is already taken.

product catalog — the *short list* of the so-called "*best of*" under the MiFID rules (Esma Guidelines on certain aspects of the MiFID II suitability requirements 2018).

The relevant research question this paper aims to answer is if a mutual fund rating based on a cardinal filter is better than a rating merely based on ordinal values. Commonly, mutual fund ratings are assigned by ranking funds according to a given performance measure (ordinal criterion), irrespective of whether a fund actually outperformed the benchmark. As a consequence, top ratings might be assigned to funds that performed poorly relative to the benchmark, thus leading to a suboptimal selection. The ordinal criterion might be "adjusted" using a cardinal filter (i.e. only fund that outperformed the benchmark can get better ratings). As for the research strategy, a new rating system based on a cardinal filter is proposed and compared to a rating system based on the ordinal criterion. Out-of-sample tests are conducted to check which of the systems performs better. The empirical analysis confirms that a rating system based on a cardinal filter performs better than the ordinal one.[b]

The paper is structured as follows: in Sec. 2, we show how the Sharpe Ratio can be used to rank mutual funds; in Sec. 3, we illustrate other risk-adjusted performance measures deriving from the Sharpe Index and sharing the same financial meaning; Sec. 4 introduces the main rating systems and the pitfalls associated with the standard approach used to assign ratings to asset management products; Sec. 5 discusses the problems arising when only ordinal measures are used, disregarding cardinal ones; Sec. 6 reports an empirical test referring to funds distributed on the Italian market, belonging to the main asset classes; in Sec. 7, an empirical test is carried out at portfolio level; Sec. 8 is devoted to conclusions.

2. The Archetype of Risk-Adjusted Performance Measures

Markowitz's (1952) model is based on the assumption that investors preferences are limited to the assessment of return and risk, and that the portfolios lying on the efficient frontier are, consistently with the mean-variance principle, the best ones. As it is known, it is not possible to rank efficient portfolios, unless a risk aversion coefficient is defined. The novelty introduced by Sharpe's CAPM (1964) consists in marking as efficient those portfolios — obtained as a combination of the market portfolio and the risk-free asset — lying on the Capital Market Line, i.e. on the line tangent to the efficient frontier with intercept the return on the risk-free asset. According to this approach, the efficiency of the market portfolio is measured by the

[b]It is out the scope of this paper to assess how the performance meaurement can affect the investment behavior as made clear by the extensive literature (Brown *et al.* 1996) on mutual funds "tournaments" and the effects of portfolio adjustments driven by the current position of a fund in a given ranking. In the paper, we assume that the number of different measures and ranking is large enough to exclude that fund managers are incentivated to game a specific one and that active managers should have always an incentive to outperform the index regardless the ranking system adopted.

slope of the Capital Market Line:

$$\frac{E(R)_{\text{Ptf}} - E(R)_{\text{Risk-free}}}{\sigma_{\text{Ptf}}}.$$

This reward-to-variability ratio (Sharpe ratio) is widely applied to measure the performance of mutual funds using time series:

$$\frac{\bar{R}_{\text{Ptf}} - \bar{R}_{\text{Risk-free}}}{\sigma_{\text{Ptf}}}.$$

This index expresses the historical excess return (over the risk-free rate) that a fund was able to create for each unit of volatility. Thus, a Sharpe index of 0.25 indicates that the fund has created 25 basis points of return (in addition to the risk-free rate) for each unit of standard deviation taken by the fund manager.

However, this Sharpe index interpretation, based on the cardinal value of the measure, does not allow a complete judgement on the efficiency of the fund, since the value assumed by the index is strongly affected by both the time window and the reference market. Table 1 shows the analysis available on the Quantalys platform[c] for three mutual funds belonging to different categories/markets that, over the three-year time window 2019–2021, recorded the same Sharpe Index — equal to 0.25 — which, however, corresponds to a different ranking in each peer group (5th quintile = very bad, 3rd quintile = average and 1st quintile = excellent).

Similarly, the same value of the Sharpe index computed in different time windows can be rated very positively in a bearish phase and very negatively in a bullish phase. What has been said so far suggests that a cardinal approach is useful to measure the efficiency, while an ordinal one is valuable to rank the fund in different *peer groups* or time windows.

The Sharpe index can (theoretically) assume values in the interval $[-\infty; +\infty]$ and, therefore, the pure cardinal measure does not allow to appreciate how difficult was to get a given performance. At the same time, a merely ordinal measure — based on the principle that a higher Sharpe Index always signals greater efficiency — can

Table 1. Sharpe index and positioning in the quintiles of the category.

FUND A			FUND B			FUND C		
3 Y measures	Relative to category		3 Y measures	Relative to category		3 Y measures	Relative to category	
Sharpe	0,25	Worst	Sharpe	0,25	Medium	Sharpe	0,25	Best
Information ratio	-1,36	Worst	Information ratio	-0,14	Medium	Information ratio	-1,08	Worst
Sortino	0,32	Worst	Sortino	0,37	Medium	Sortino	0,35	Best
Omega	1,17	Worst	Omega	1,10	Medium	Omega	1,11	High
Calmar ratio	0,13	Worst	Calmar ratio	0,11	Low	Calmar ratio	0,18	Best

[c] https://quantalys.it/.

be misleading, as it is unable to capture both the different significance of a negative rather than a positive value and the gap between the funds ranked.

A negative Sharpe Ratio expresses the basis points (in excess of the return on the risk-free asset) "destroyed" per unit of risk, and therefore, a higher value of this indicator signals greater efficiency (in this case interpreted as less destruction of value). However, the negative value of the Sharpe's Index seems to be totally counterintuitive to the common feeling of investors, who would hardly be willing to prize mutual funds that have experienced huge losses with high levels of risk, rather than funds showing limited losses with low levels of risk. An example may be useful to understand the paradox attributable to a negative value of the Sharpe index. Assume that fund A recorded a 1% loss with a 2% risk and fund B recorded a 5% loss with a 10% risk. Assuming a risk-free rate of 1%, fund A has a Sharpe ratio of -1 and fund B a value of -0.6. On the basis of an ordinal valuation, fund B is, therefore, more efficient. Assuming that the benchmark also recorded a loss of 3% with a risk of 4%, the counterintuitiveness of the Sharpe Index is evident: it suggests to select fund B, despite the fact it recorded a lower return than the market with a higher risk, and to reject fund A despite its ability to outperform the market while reducing the risk taken. Therefore, a ranking based on the use of Sharpe's Index is unable to take into account both the cardinal value and the sign of the ratio.

Equally misleading is an ordinal valuation of funds that fails to take due account of the difference in terms of cardinal value. Table 2 shows three cases where despite the presence of significant differences in terms of cardinal values, the ordinal values turn out to be equal, thus producing the same ranking of funds with completely different measures of risk-adjusted performance. In other words, the same ranking on the ordinal scale is recorded regardless of the presence of significantly different Sharpe Index values both in absolute terms and in terms of distance from the *best performer*.

Table 2. Mean ordinal values in the presence of heterogeneous cardinal values.

Case A		Case B		Case C	
Cardinal values	Ordinal values	Cardinal values	Ordinal values	Cardinal values	Ordinal values
+0.750	1	+0.750	1	−0.60	1
+0.749	2	+0.700	2	−0.65	2
+0.748	3	+0.650	3	−0.70	3
+0.747	4	+0.600	4	−0.75	4
+0.746	5	+0.550	5	−0.80	5
+0.745	6	+0.500	6	−0.85	6
+0.744	7	+0.450	7	−0.90	7
+0.743	8	+0.400	8	−0.95	8
+0.742	9	+0.350	9	−1.00	9
+0.741	10	+0.300	10	−1.05	10

3. Alternative Risk-Adjusted Performance Measures

The Sharpe Index, despite its popularity and notoriety, has some limitations that have encouraged the use of alternative measures. The following is a review of the main measures with the aim of highlighting how they differ from the Sharpe Index.

Modigliani's RAP (*risk-adjusted performance*) index (Modigliani & Modigliani 1997) aims to measure the return that the portfolio would have obtained assuming a level of risk equal to that of the benchmark (M). For the purpose of quantifying the Modigliani Index, it is, therefore, necessary to build a virtual portfolio with risk equal to that of the benchmark; consistently with the CAPM hypothesis, given a portfolio of *risky assets*, it is possible to change the risk profile by investing or borrowing in the risk-free asset:

$$I_{\text{Modigliani}} = \bar{R}_{\text{Risk-free}} + \text{Sharpe Index} \times \sigma_{\text{Benchmark}}.$$

Since the Modigliani Index of a panel of funds depends on two constants (*risk free* rate and standard deviation of the *benchmark*) and on the Sharpe Index of each fund, Sharpe ratio and Modigliani index show different values for each fund (cardinal values of efficiency), but they lead to the same fund ranking.

Although Sharpe's Index and Modigliani's Index measure efficiency in different ways, they capture risk with the same measure — the standard deviation — which, however, is unable to take into account the presence of skewness or kurtosis phenomena in the return distribution and the consequent need to penalize funds showing *left-tail risk*. Furthermore, standard deviation defines risk as the dispersion of returns around the mean, without any discrimination between returns above or below the average.[d]

Since the returns of a portfolio can assume a non-Gaussian distribution,[e] it is useful to use volatility measures that can penalize funds characterized by negative skewness and fat left tails. Moreover, it makes sense to assume that investors, rather than being averse to the generic dispersion of returns (to the standard deviation), are concerned about the occurrence of negative excess returns located on the left-hand side of the return distribution. Measures attempting to capture the behavior of this (downside) portion of the distribution are called lower partial moments (lpm). Assuming that the investor is only averse to returns below a threshold value h, returns above the threshold should be ignored, not being a risk. Often the threshold h is set equal to the risk-free rate and the gray portion of the return distribution shown in Fig. 1 identifies the observations that should be used in estimating the partial

[d]The first to raise doubts about the use of standard deviation as a risk parameter was Markowitz (1959). The use of the standard deviation was a choice "of convenience", suggested by the need not to excessively increase complexity of the optimization model: "*One of the measures considered, the semideviation, produces efficient portfolios some what preferable to those of the standard deviation. Those produced by the standard deviation are satisfactory, however, and the standard deviation itself is easier to use, more familiar to many, and perhaps easier to interpret than the semideviation* (...) *Efficient portfolios based on variance, however, cannot be characterised as bad or undesirable*".

[e]See Singleton & Wingender (1986), Peiro (1994, 1999).

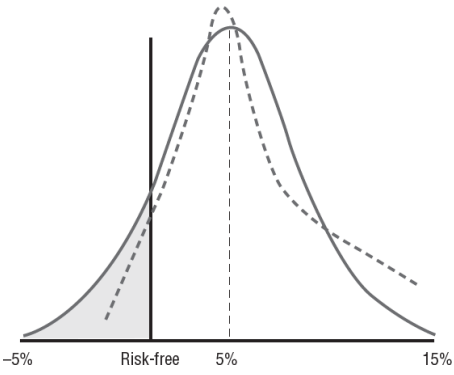

Fig. 1. Standard deviation, downside risk and return distribution.

volatility that investors would like to minimize. Such measures of partial volatility don't penalize the asset manager for being able to amplify volatility during the "bull" phases.

The *downside risk* (Sortino & van der Meer 1991) represents a measure of the dispersion of returns below a threshold — which can be interpreted as the minimum acceptable return (*Minimum Acceptable Return*, h) — normally identified in the risk-free rate or the zero return. *Downside risk* is calculated as follows:

$$\mathrm{DSR} = \sqrt{\sum_{i=1}^{N} \frac{(R_i - h)^2 d_i}{N}},$$

with :

$$d_i = \begin{cases} 0 & \text{if } R_i \geq h, \\ 1 & \text{if } R_i < h. \end{cases}$$

This measure is complementary to standard deviation since it enriches its information by giving special emphasis on returns below the threshold. Similarly, the Sortino Index (Sortino & Satchell 2001), showing at the denominator the *downside risk*, provides additional information on the efficiency of the fund. This measure is particularly appropriate for those types of products (flexible funds, absolute return funds, *hedge funds*, etc.) that make loss minimization the key element of their management style. In formulas,

$$\text{Sortino Index} = \frac{\bar{R}_{\mathrm{Ptf}} - \bar{R}_{\text{Risk-free}}}{\mathrm{DSR}_{\mathrm{Ptf}}}.$$

The family of asymmetric risk measures includes other *lower partial moments*:

– zero-order *lpm* identifies the shortfall probability, i.e. the probability that returns below threshold h will occur;
– first-order *lpm* measures the average *shortfall*, better known as *expected shortfall* or *Conditional VaR*.

Among the most widely used asymmetrical measures, the *maximum drawdown* (MAX DD) is worthy of mention; it quantifies the maximum loss incurred by a portfolio within a given time window. The peculiarity of *maximum drawdown*, which distinguishes this indicator from all the previous ones, is its ability to take into account the time sequence of returns. In fact, while measures such as standard deviation and *downside risk*, if calculated through a simple moving average logic, do not see their value varying by modifying the order of returns in the time series, *maximum drawdown is significantly influenced by the sequence of returns.*

Similar to what has been observed for *downside risk*, the *maximum drawdown* can also be used to calculate a *risk-adjusted* performance indicator. By replacing in Sharpe's Index formula *the* standard deviation with the MAX DD, the Calmar Ratio is built (Young 1991). In formulas,[f]

$$\text{Calmar ratio} = \frac{\bar{R}_{\text{Ptf}} - \bar{R}_{\text{Risk-free}}}{\text{MAX DD}_{\text{Ptf}}}.$$

In order to measure active management, risk can be carried out from a further perspective that examines the time series of differential returns with respect to a *benchmark* consistent with the category to which the fund belongs. This is a risk quantification that is entirely consistent with the process leading to the fund selection, starting from the identification of the client's investment objectives and his risk tolerance, then moving on to the definition of the asset allocation in market indices to conclude with the translation into products. In fact, the risk contribution coming from the choice of active managers cannot be quantified by the previous indicators, since they are affected both by the market behavior and by the activism of the fund managers. Relative risk measures neutralize the contribution of the market factor and offer a representation of the risk that makes it possible to discriminate among managers according to their different sensitivity to market fluctuations (β) or to the magnitude of the volatility of the bets made against the benchmark (*tracking error volatility*).

The beta (β) of the portfolio is a measure derived from the Capital Asset Pricing Model; graphically, it is the slope of the regression line obtained by regressing the portfolio's *excess returns* (dependent variable) with those of the market (independent variable). A beta equal to 1 identifies a neutrality of the portfolio with respect to market variations (which could also derive from a passive management style), while values above or below 1 identify an active management style (aggressive or defensive). As for other measures of risk, beta can be used to obtain an alternative (but scarcely used) *risk-adjusted* performance measure named Treynor's Index (Treynor 1965):

$$\text{Treynor Index} = \frac{\bar{R}_{\text{Ptf}} - \bar{R}_{\text{Risk-free}}}{\beta_{\text{Ptf}}}.$$

[f]Although the original version proposed by Young uses the absolute return in the numerator, in practice, the excess return over the risk-free rate is commonly used.

The tracking error volatility (Sharpe 1994, Tobe 1999), on the other hand, is an indicator of relative risk focused on the bets made by the asset manager against the benchmark; therefore, this parameter has values close to 0 for passive funds and gradually increasing values as the degree of "activism" of the asset manager increases. In formulas,

$$\text{TEV} = \sqrt{\frac{\sum_{i=1}^{N} (\text{TE}_i - \overline{\text{TE}})^2}{N}}.$$

A high tracking error volatility signals an active management — which is not necessarily linked with a higher risk than the benchmark — but is not able to provide any information on the goodness of investment choices (Sharpe 1991). The risk-adjusted return measure capable of combining information on the efficiency (return per unit of risk) and effectiveness (ability to beat the benchmark) is represented by the Information Ratio (Sharpe 1994) which places in the numerator the portfolio's *excess return* with respect to the benchmark — thus removing any reference to risk-free rate — and in the denominator the tracking error volatility, i.e. the degree of "infidelity" of the portfolio to the benchmark:

$$\text{Information ratio} = \frac{\bar{R}_{\text{Ptf}} - \bar{R}_{\text{Mkt}}}{\text{TEV}_{\text{Ptf}}}.$$

In conclusion, the use of different measures of excess return (risk premium or tracking error) and risk (standard deviation, downside risk, maximum drawdown, beta or tracking error volatility) adopted in the risk-adjusted performance calculation, do not allow to identify the best measure unless the objective of the analysis is not clearly defined.

4. Measures Used by Rating Companies

The review of risk-adjusted return measures in the previous section cannot be assumed to be exhaustive of the multiplicity of indicators suggested by academic research and market practice. In particular, it is worth mentioning the measures used by fund rating companies which assign a score expressed in the number of stars varying from 1 to 5. The companies analyzed below, by virtue of their wide visibility on the Italian market, are Morningtar, Quantalys and Lipper.

Morningtsar, a world leading US company in the valuation of asset management products, attributes the quantitative rating via a risk-adjusted return indicator capable of attributing different weights to the downside deviation than to the upside deviation. Morningstar's proprietary indicator[g] (Morningstar Risk-Adjusted Return — MRAR) is based on the expected utility theory, by virtue of which an investor ranks the various investment alternatives in his or her portfolio on the basis of the

[g]See, https://www.morningstar.com/content/dam/marketing/shared/research/methodology/771945_Morningstar_Rating_for_Funds_Methodology.pdf.

final value of that investment at the end of a given time window (3 years, 5 years, 10 years, etc.). In defining this value, Morningstar uses excess returns adjusted by the impact of fund entry fees.

The algorithm behind the Morningstar approach is based, similar to the other indicators analyzed so far, on the use of time series returns; therefore, this measure must also be placed in the family of *backward-looking* measures. In formulas,

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

where γ represents the investor's risk aversion, which in the Morningstar model takes a value of 2; ER identifies the fund's excess returns over the risk-free rate adjusted for entry fees; T is the time window (expressed in months).

Quantalys[h] rating — as well expressed in terms of stars — is based on a multivariate methodology that is a joint evaluation of three elements: (a) the ability to beat the benchmark of the category to which the fund belongs (which is therefore the same for all investment products sharing the same peer group); (b) the ability to beat the style benchmark (which differs from fund to fund); (c) the ability to anticipate bullish and bearish phases of the markets. The combination of these three factors, appropriately normalized, returns an overall assessment of the portfolio's performance, which is then used to rank the products belonging to the same peer group.

Lipper[i] proposes a plurality of ratings that refer: (i) to absolute return without any risk adjustment, (ii) to the systematic of the return achieved based on a risk-adjusted return measure and (iii) to capital preservation measured as the sum of negative monthly returns over the time window.

5. Cardinal Versus Ordinal Measures

Each of the risk-adjusted return measures analyzed in Sec. 3 can be used to evaluate mutual funds using either a cardinal or an ordinal ranking. For example, in the case of the Information Ratio, the adoption of a cardinal or ordinal ranking may lead to different conclusions. In fact, it is well known that the sign of the information ratio (IR) makes it possible to discriminate between the funds capable of beating the market ($\mathrm{IR} > 0$) from those that underperformed it ($\mathrm{IR} < 0$). While the value of the Information Ratio can signal the greater or lower efficiency of the asset manager in terms of basis points added or destroyed for a unit of tracking error volatility.

Which are the paradoxes arising from the translation of a risk-adjusted return indicator into a synthetic judgment represented by stars? To answer this question, it is worth pointing out that the universal approach adopted to rate funds is: (1) to analyze investment products belonging to the same peer group, (2) to rank them

[h] See, https://www.quantalys.it/Whoarewe/Methodology.
[i] See, https://lipperalpha.refinitiv.com/wp-content/uploads/2016/01/RE1260174_IA_factsheet_A4_Lipper-Leaders-Intro.pdf.

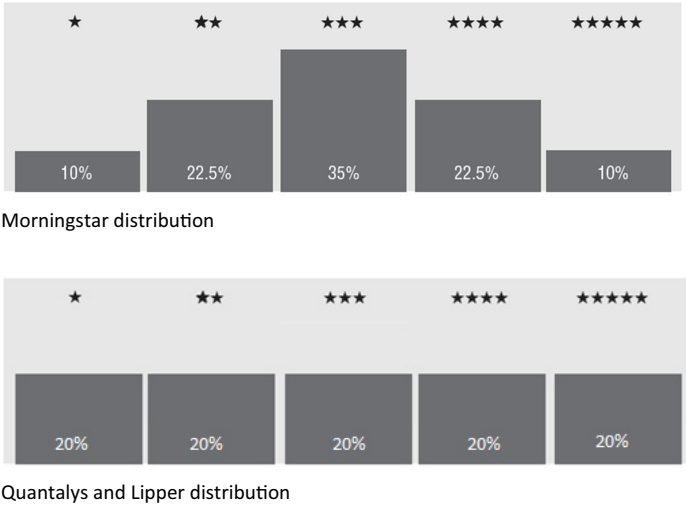

Fig. 2. Distribution of ordinal measures and rating assignment.

according to a specific risk-adjusted measure and (3) to assign stars on the basis of a statistical distribution of the population. Figure 2 shows a comparison of the distributions performed by Morningstar, Quantalys and Lipper.

The abovementioned methods of rating assignment completely disregard the signaling value of the cardinal measure of the chosen indicators; therefore, mutual funds are evaluated by giving more importance to their position in the peer group than to the size of the manager's skill or efficiency. As a consequence, it may happen that a fund with a positive Information Ratio (and thus able to fulfill its promise of beating the market) may receive a low rating — 1 or 2 stars — simply because it is characterized by a risk-adjusted performance measure even slightly lower than that of funds ranked in the first classes. Certainly, more worrying is the circumstance that such an approach can assign a brilliant rating — 4 or 5 stars — to a fund with a negative IR (unable to fulfill its promise of beating the market) simply because it has a higher risk-adjusted performance measure than its competitors. Based on what has just been stated, there is a clear risk that the rating of excellence is awarded to asset managers who do not deserve it. Assuming a normal distribution of funds ranked on the basis of a given risk-adjusted return indicator, a shift — i.e. a change in the mean value — to the right or left (Fig. 3) would in no way alter the rating system.

As a second limitation of standard rating systems based on ordinal measures, they distribute products within a limited number of groups that are not homogeneous within themselves. That is, there is the risk that products having received the same number of stars present a high volatility of the value of the indicator chosen for the classification.

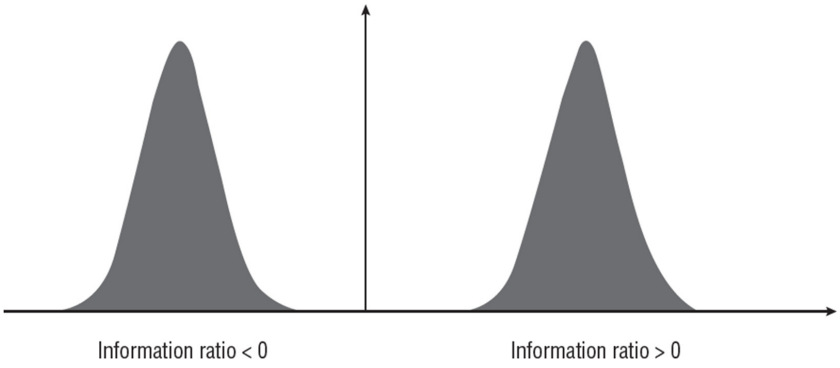

Fig. 3. The distribution of the Information Ratio: absolute and relative values.

6. Empirical Analysis (Data and Methodology)

In order to answer the research question underlying this paper (*can a judgement of excellence of an asset management product be derived from a ranking drawn up using ordinal values or is it preferable to introduce a filter based on cardinal values?*), a sample was chosen made up of all the "blend" funds (included in the Quantalys database) distributed on the Italian market and managed by domestic and foreign fund companies. The sample (Table 3) covers the main categories of equity (Europe, North America, Japan and Emerging Countries) and bond (Euro, Global and Emerging Countries) products. Funds with more than one class were considered once, selecting only the main class; all performances were calculated in Euro and are total return (for funds with distributed income, dividends were reinvested in the same fund). Finally, survivorship bias was taken into account.

The sample consists of 1838 funds, analyzed over the time window from 30 March 2001 to 2 June 2022. Each category was associated with a market index, whose performance was adjusted to take into account an investment cost of 20 basis points per year.

Table 3. Composition of the sample by categories.

Category/Asset class	Number of funds	Reference index
(1) Europe equity	312	MSCI Europe TR
(2) Noth America equity	221	MSCI North America TR
(3) Japan equity	246	MSCI Japan TR
(4) Emerging Markets equity	136	MSCI EM TR
(5) Euro bond	217	ICE BofA All Euro Government Index — TR
(6) Global bond	405	ICE BofA Global Government Index — TR
(7) Emerging Markets bond	301	ICE BofA Diversified Emerging Markets External Debt Sovereign Bond Index TR

The first objective of the empirical test was to verify the impact caused by the change in rating criteria moving from a purely ordinal to a cardinal approach. More specifically, by ordinal methodology, we mean: (a) the attribution of the rating by means of a given risk-adjusted return indicator, (b) the sorting of the values of the risk-adjusted performance measures and (c) the final identification of the percentiles by which stars are attributed (5 to excellent funds and 1 to mediocre funds[j]). The cardinal methodology, on the other hand — in line with the literature on active management (aimed at generating net returns above market returns) and passive management (aimed at replicating the net returns of the reference benchmark) — aims at rewarding managers who are able to achieve their objectives and not simply those who are better ranked than the others. The assumption of the cardinal methodology is that only active funds that, thanks to *stock picking* and *market timing*, have been able *to* beat the market and passive funds characterized by management costs below the market average (and thus with returns above those of the benchmark) can be awarded a 3-star or higher rating.

Consistent with this approach, the first step was to assess the impact of the criterion adopted (ordinal *versus* cardinal) on the rating attribution to the funds belonging to the seven aforementioned categories. We use a three-year (156-week) time window ranging from 31/05/2019 to 02/06/2022. As the rating attribution methodology was applied to a plurality of *risk-adjusted performance measures*, the analysis is *measure independent*. Among the available indicators, we have selected the Information Ratio, since it is the only measure able to consider the manager's skill and efficiency.[k]

Specifically, with reference to the ordinal criterion, a ranking was drawn up on the basis of the abovementioned risk-adjusted performance measure, and a rating of 5 stars was attributed to the funds belonging to the first quintile, up to 1 star for the funds belonging to the fifth quintile; in this way, the funds are uniformly distributed across the five rating classes.

As far as the cardinal criterion is concerned, the ranking of the funds was "cut" at a null value of the tracking error and the risk-adjusted return measures were then normalized[l] in a range [−100; +100], attributing a positive value [0; +100] to the "best" funds (which beat the market) and a negative value [−100; 0] to the "worst" funds (which underperform the market). The rating attribution on the basis of the

[j] For the purposes of this work, a homogeneous distribution of funds was made, assigning 5 stars to funds in the first quintile, 4 stars to those in the second quintile, 3 stars to funds in the third quintile, 2 stars to funds in the penultimate quintile and 1 star to funds in the last quintile. The percentile distribution made by Morningstar gives a rating of 3 stars or more to 67.50% of the funds and, therefore, amplifies the differences between ordinal and cardinal criteria.

[k] We also conducted the same empirical analysis using the Sharpe and the Sortino Ratios and the results are not significantly different.

[l] A linear normalization was used, assigning a zero score to the worst fund being part of the subset of funds beating the benchmark and 100 score to the best fund; similarly, at the bottom of the ranking (the subset of funds underperforming the benchmark), a value of 0 is assigned to the best fund and −100 to the worst fund. There are no evaluation problems related to the presence of two funds both having a zero score, as they belong to two distinct clusters.

Table 4. Rating with the cardinal criterion.

Funds in the peer group	Range of the normalized measure	Rating
Funds with non-negative information ratio - Best Funds	$66.6 < X \leq 100$	*****
	$33.3 < X \leq 66.6$	****
	$0 \leq X \leq 33.3$	***
Funds with negative information ratio - Worst Funds	$-50 < X \leq 0$	**
	$-100 < X \leq -50$	*

values thus obtained is based on the rule that a higher score must correspond to a better rating attributed on the basis of the criterion represented in Table 4: the funds are not uniformly distributed across the five risk classes, as the distribution is conditioned by the absolute value assumed by the risk-adjusted return measures.

The comparison between the ordinal and cardinal rating assignment

– makes it possible to quantify the percentage of funds that receive a positive rating (rating ≥ 3) with the ordinal criterion despite having achieved a performance that is not in line with the objectives to exceed the benchmark return (in the case of active funds) or to replicate it with low costs (in the case of passive funds);
– leads to a transition matrix that makes it possible to verify if, given the same measure of risk-adjusted performance, the judgement expressed on a fund may be model dependent, i.e. influenced by the different criterion — ordinal or cardinal — used to assign the rating.

In relation to the first point, we observe (Fig. 4) that the number of funds that received a positive rating (equal to or higher than 3) despite having underperformed the market is 56% for equity funds and 12% for bond funds.

The empirical analysis makes it possible to state that a positive rating given to funds that do not deserve it constitutes a significant phenomenon to which the ordinal criterion is exposed; this problem, even affecting funds that received 5 stars, emerges with greater evidence for equity funds and, in particular, for asset classes characterized by a higher level of efficiency: the North American equity asset class shows the highest percentage of funds receiving a positive rating (≥ 3) despite having underperformed the market. As it is known, the greater the efficiency the harder it is to generate extra performance through *stock picking* and *market timing* and, consequently, the higher the probability of underperforming the market (also due to higher costs). Thus, if absurdly all funds of a category performed worse than the market — unequivocally signaling the absolute convenience of using passive management — the ordinal rating criterion would in any case assign 60% of the funds a rating of 3 stars or higher.

The percentage of funds with a return below the benchmark is, in contrast, significantly lower for bond funds, and in particular, for the Euro Bond asset class where only 6% of funds with a rating equal to or greater than 3 stars performed worse

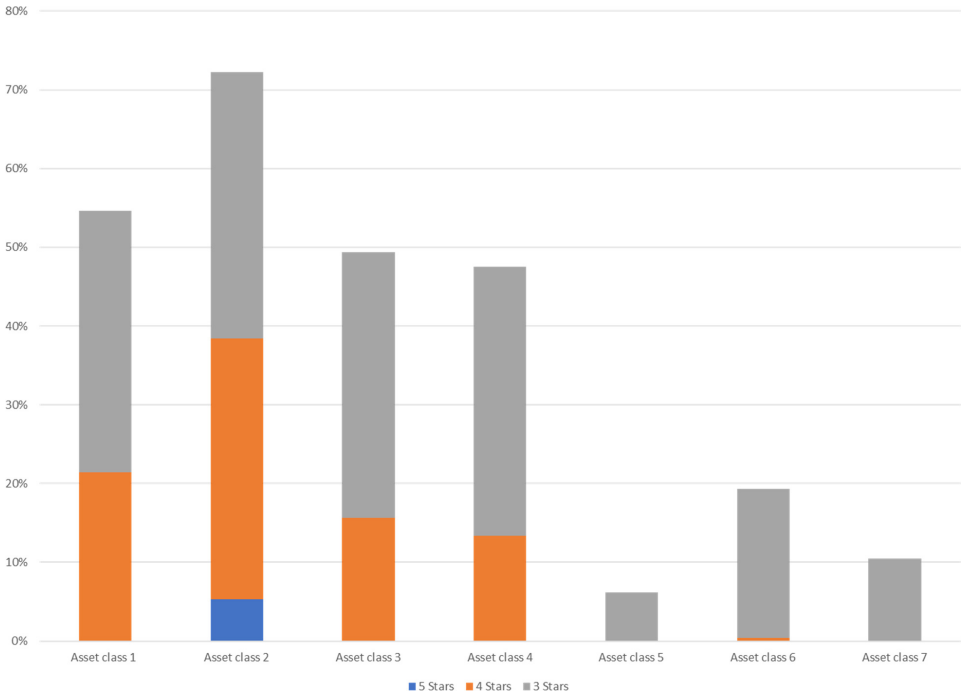

Fig. 4. Percentage of funds with a 3-star or higher rating (ordinal Information Ratio criterion) and below-market performance.

than the market. This is in some ways surprising, given that this asset class is also characterized by high levels of efficiency, being dominated by the weight of government bonds. A plausible explanation is that in the time window analyzed (31/05/2019 to 02/06/2022) most managers were able to make portfolio choices capable of effectively managing the presence of negative interest rates.

The percentages do not change appreciably when we look at the other risk-adjusted performance measures (Sharpe Index and Sortino Index): the only exception concerns the Euro bond asset class, for which we found a marked increase in the number of funds underperforming the market with ratings ≥ 3 stars. Comparing the percentages recorded with different *risk-adjusted* return measures, we conclude that the problem of award-winning funds that don't deserve it is more affected by the reference market (and its efficiency) than by the *risk-adjusted* return indicator used to assign the rating.

The empirical analysis also makes it possible to verify whether the use of the ordinal rather than the cardinal criterion produces significant effects on the rating class assignment. If, in fact, the two criteria led to the same output, the cardinal criterion would be useless, by virtue of its inability to promote changes in the rating assignment. This "extreme" hypothesis is described graphically in Fig. 5 in which the x-axis shows the ratings assigned with an ordinal logic and the y-axis shows those

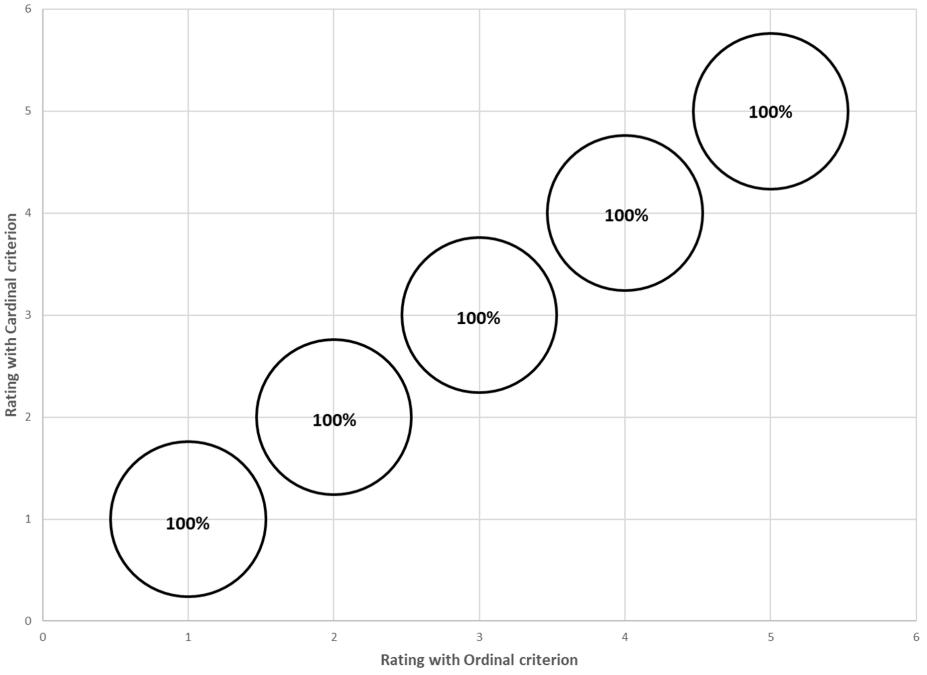

Fig. 5. Migration between rating classes assuming overlap between ordinal and cardinal criteria.

assigned with the cardinal criterion; the size of the bubbles measures the percentage of funds retaining the same rating, which, under the hypothesis of neutrality of the cardinal criterion, is 100%.

Empirical evidence shows, however, that

– rating shifts are significant and, for the best ratings, almost always worsening;
– the shift from ordinal to cardinal criteria reduces the number of funds with better ratings, drastically limiting the number of funds labeled with a rating ≥ 3 stars;
– the set of funds awarded by the rating analysis shows a reduction that averages 55.90% for equity and 12% for bond funds.

In order to simplify a comparison of the results obtained for the seven asset classes, Table 5 summarizes the number of funds and the relative percentage that recorded a change in the rating received.

It emerges that the cardinal criterion generates a significant impact on the rating assignment; as confirmed by the following figures, this phenomenon is pronounced for funds that received (with the ordinal criterion) a rating of 3 stars or more.

Specifically, for the European equity (Fig. 6), only 19.4% of the funds rated 5 under the ordinal criterion retain the same rating; for funds with 3 and 4 stars, all of them receive a worse rating than the ordinal one; funds with 1 or 2 stars retain a negative rating (<3), although there is a 1 notch upgrade for the worst funds.

Table 5. Funds affected by rating change.

Asset class	Number of funds	Number of migration	% migrated funds
Asset class 1	312	232	74%
Asset class 2	221	151	68%
Asset class 3	246	172	70%
Asset class 4	136	86	63%
Asset class 5	217	108	50%
Asset class 6	405	235	58%
Asset class 7	301	182	60%

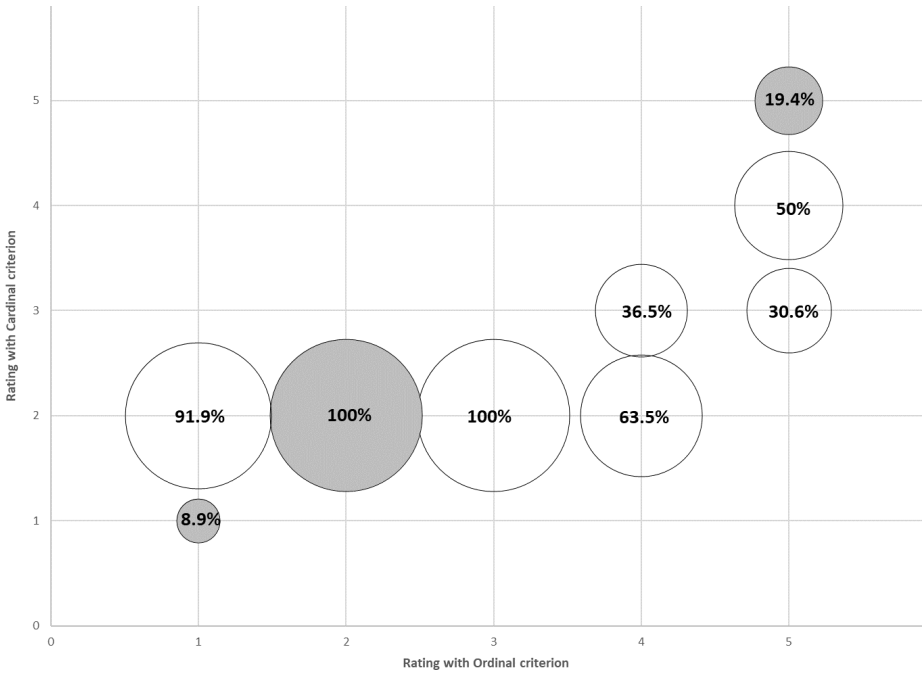

Fig. 6. Migration among rating classes of Europe equity funds.

Moving on to the North American equity (Fig. 7), only 11.4% of funds rated 5 or higher with the ordinal criteria retained the same rating, and as many as 15.9% lost three notches, receiving a 2-star cardinal rating; in addition, 100% of funds that received 4 and 3 stars with the ordinal criteria were downgraded by 2 and 1 notch, respectively, receiving a 2-star cardinal rating. The high number of downgrading shown by funds rated 3 or more stars via an ordinal basis is fully consistent with the high efficiency of the North American equity market and the consequent difficulty of overperforming the market. Finally, for funds with an ordinal rating < 3 stars, changes observed are not significant, apart from an almost physiological 1 notch upgrade of ordinal 1-star funds.

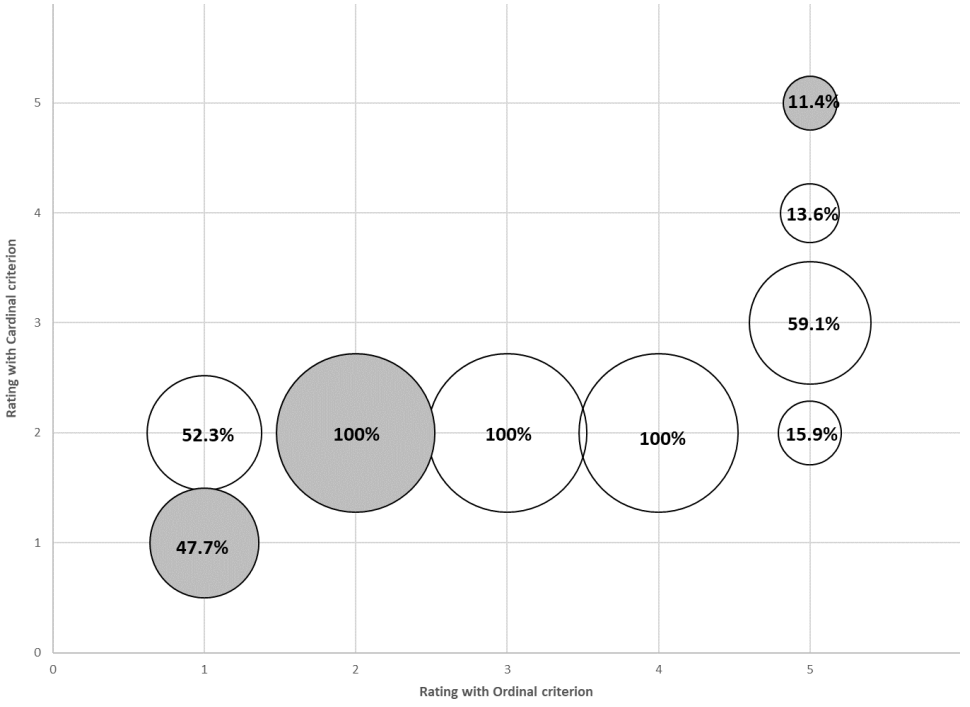

Fig. 7. Migration among rating classes of North America equity funds.

For the Emerging Markets equity (Fig. 8), there was also widespread downgrading of funds with ordinal rating of 3 stars or more. Only 4.8% of funds with an ordinal 5 stars rating retained this class; for all the other funds, downgrading was 1 or 2 notches. In the case of funds with an ordinal of less than 3 rating, we observed

– a perfect persistence in the case of a rating of 2;
– an upgrade of 1 notch for 53.1% of the 1-star funds.

Japan equity (Fig. 9) shows results that are in line with those of the Emerging Markets equity: 22% of funds with an ordinal rating of 5 stars were able to confirm this rating; this is a further confirmation that the majority of funds having an ordinal rating of 3 or higher was downgraded by 1 or 2 notches. In the case of funds with an ordinal <3 rating, we observe

– a perfect persistence in the case of a 2-star rating;
– an upgrade of 1 notch for 37% of the 1-star funds.

Turning to the bond asset classes, a higher persistence is generally observed. In the case of the Euro bond (Fig. 10), 23.3% of the funds retain a rating of 5, while 76.7% of ordinal 5-star funds and 100 % of ordinal 4-star funds recorded a 1 notch

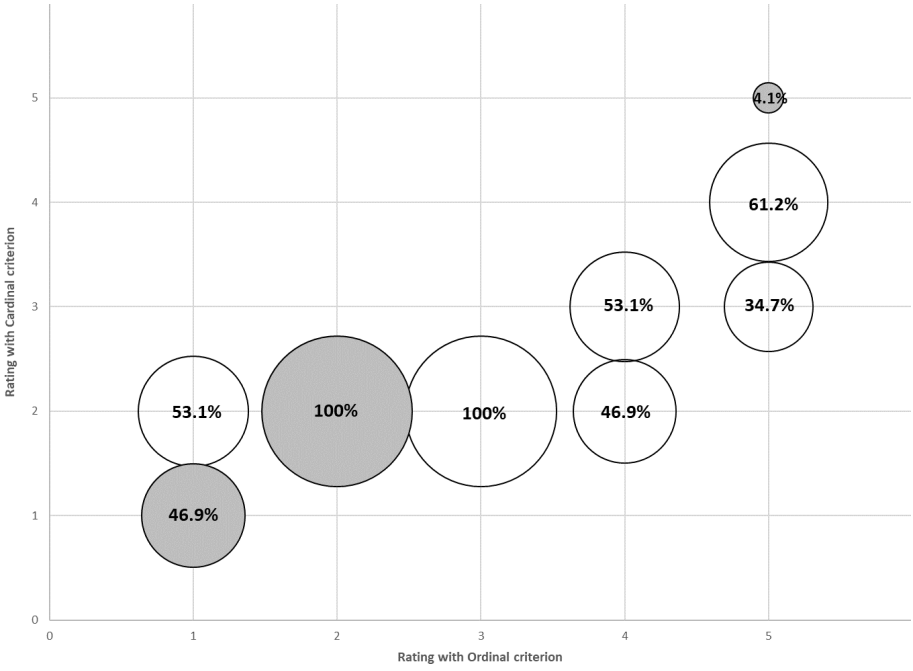

Fig. 8. Migration among rating classes of Emerging Markets equity funds.

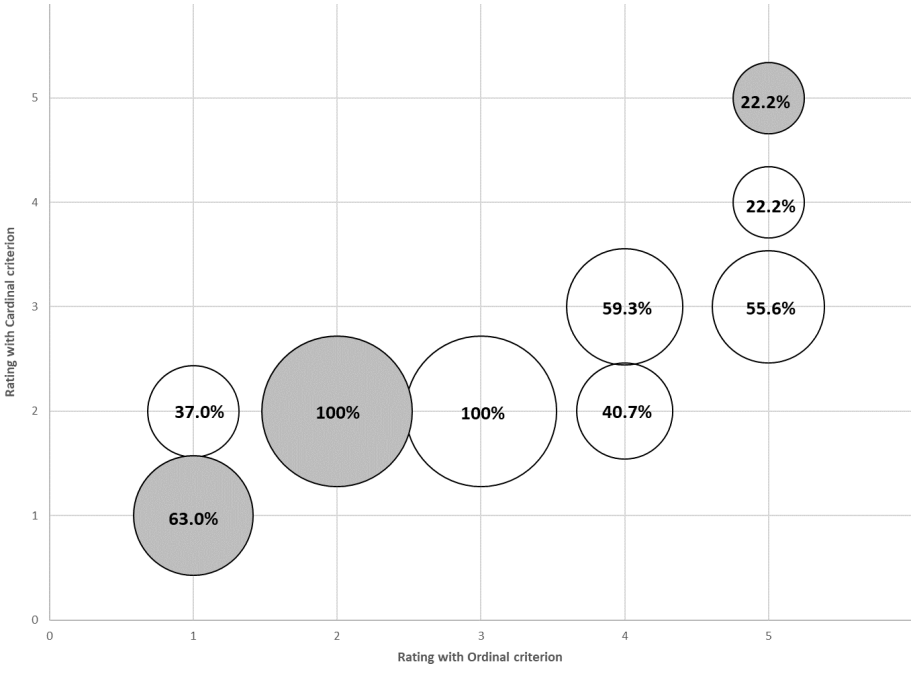

Fig. 9. Migration among rating classes of Japan equity funds.

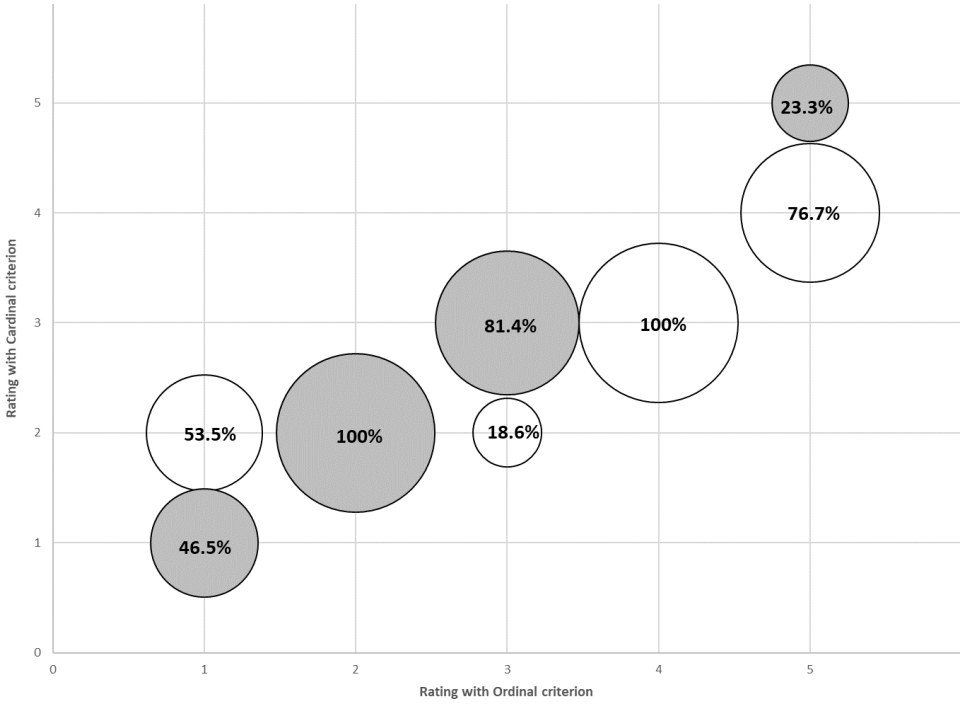

Fig. 10. Migration among rating classes of Euro bond funds.

downgrading. There is also a good persistence in the case of funds with an ordinal 3-star rating: 81.4% of them confirm the class. For funds with an ordinal rating <3, no behavior differing from that already observed for equity funds is recorded.

Similar evidence emerges from the Global bond asset class (Fig. 11), which, compared to the Euro bond asset class, records a more pronounced downgrading of funds that received a ≥ 3 ordinal rating.

The Emerging Countries bond asset class (Fig. 12) shows a poor persistence (10%) for the extreme classes (1 and 5 stars); on the contrary, persistence is absolutely higher for intermediate classes (2 and 3 stars). Migrations are only a single notch.

The empirical analysis allows us to highlight that the rating attribution is *model dependent* and, therefore, the choice of the criterion (cardinal versus ordinal) assumes a relevance that is at least as important as the choice of the *risk-adjusted performance* indicator. The research question we have posed is, therefore, worthy of attention; further investigation is required to verify which of the two methodologies performs better in predicting which funds will beat the competitors.

In order to verify the potential advantages of a cardinal rating assignment, which can also significantly limit the number of funds worthy of selection, we investigated

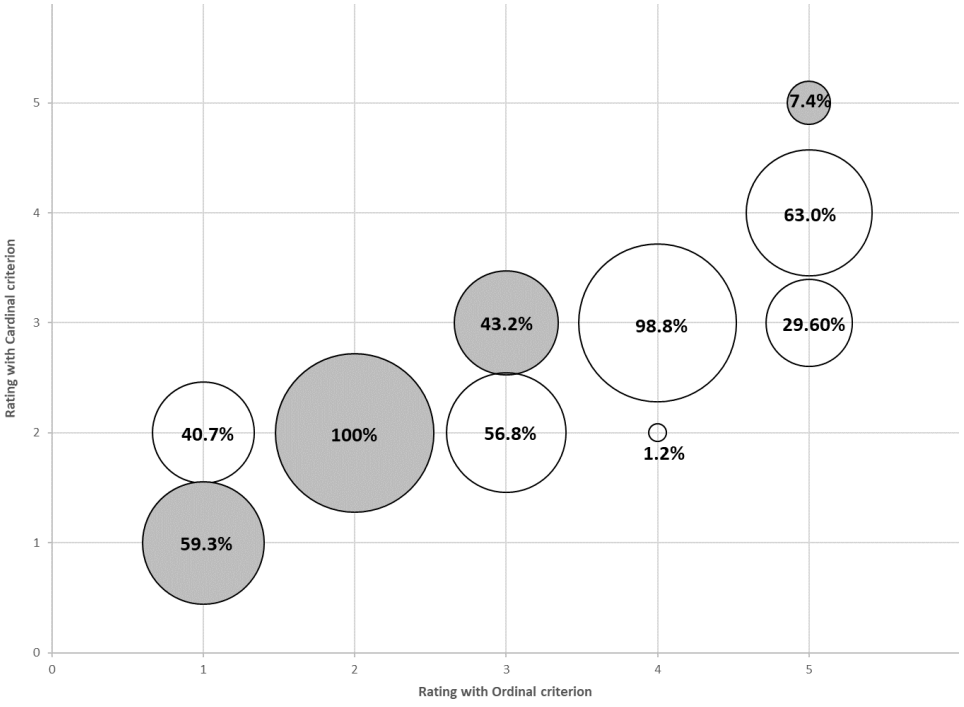

Fig. 11. Migration between rating classes of Global bond funds.

three different reliability profiles of the rating assigned with different criteria:

– over the three-year time window following the rating attribution, do the cardinal 4- and 5-star funds perform better than the ordinal 4- and 5-star funds?[m]
– does the return of a portfolio obtained by annually selecting the cardinal 4 and 5-star funds exceeds that obtained by investing in ordinal 4- and 5-star funds?
– do the funds classified in rating classes correctly anticipate the result obtained in the three years following the rating attribution?

The first empirical test was conducted starting with the three-year time window 30 March 2001–26 March 2004 and then proceeding with a 13-week rolling analysis up to the most recent three-year time window (3 June 2016–29 May 2019). The choice of the three-year window is consistent with the time horizon typically used by mutual fund rating companies; the choice of a 13-week rolling period is consistent with the quarterly review on fund catalog that banks perform to identify the short lists of best-in-class products. For each of the 61 rolling periods thus identified, an equal-weighted portfolio of 4- and 5-star funds was set up both for the ordinal and cardinal

[m]We assume that new investment flows are concentrated in funds ranked 4 and 5 stars. This evidence is confirmed by Del Guercio & Tkac (2008) which also show that investors "punish" funds that drop to a 3-star Morningstar rating.

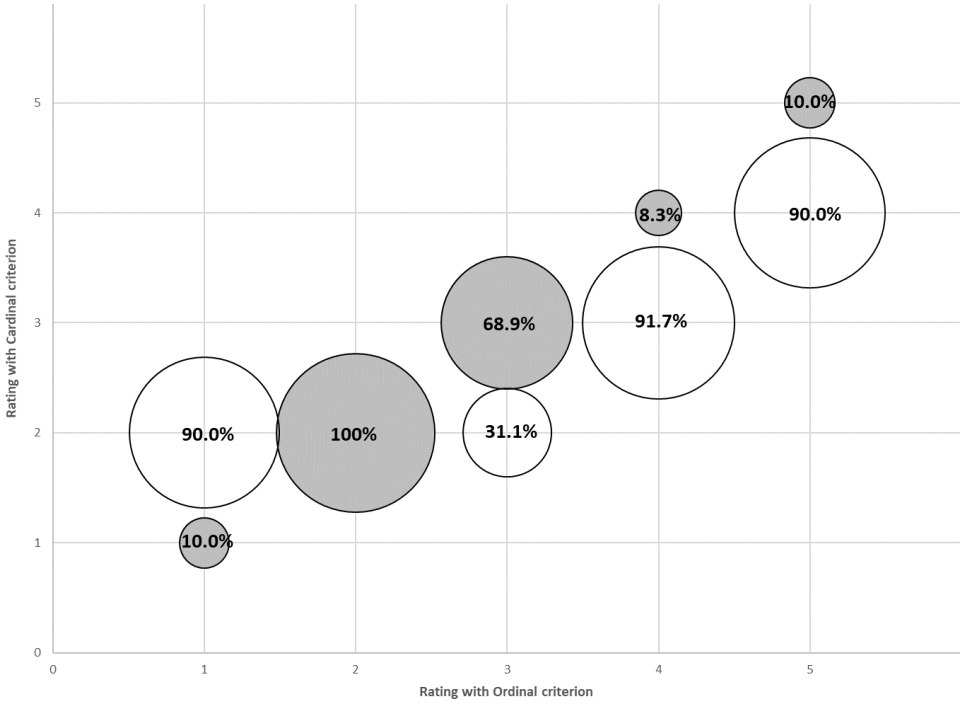

Fig. 12. Migration among rating classes of Emerging Markets bond funds.

assignments. Then, we have analyzed in how many cases the best fund portfolio consistent with the cardinal approach outperformed the other consistent with the ordinal method, in the three years following the rating assignment. The empirical evidence (Fig. 13) shows that in the majority of the 61 time windows, the portfolio selected with the cardinal approach outperformed the other one. Our analyses confirm the results are not asset class-dependent or risk-adjusted performance-dependent (the only exception concerns asset class 7 — Emerging Countries bond — in the case of the Sortino Index).

As for the comparison of the riskiness, empirical evidence has not shown, as for return, the superiority of the portfolios built with the cardinal criterion.

The second empirical test was conducted by constructing the equally weighted portfolios at the end of the first three-year period of analysis (30 March 2001–26 March 2004) composed by the funds ranked 4 or 5 stars based, respectively, on the cardinal and ordinal criteria. The composition of the two portfolios was updated — based on the same approach — using a 13-week rolling window and then assuming that a quarter represents the minimum time horizon for assessing a mutual fund's performance. In fact, working with shorter time windows, the manager's skill may be confused with luck. One year is also the minimum window for a fund to remain in the portfolio. Given the size of the time series (30 March 2001–2 June 2022), the fund

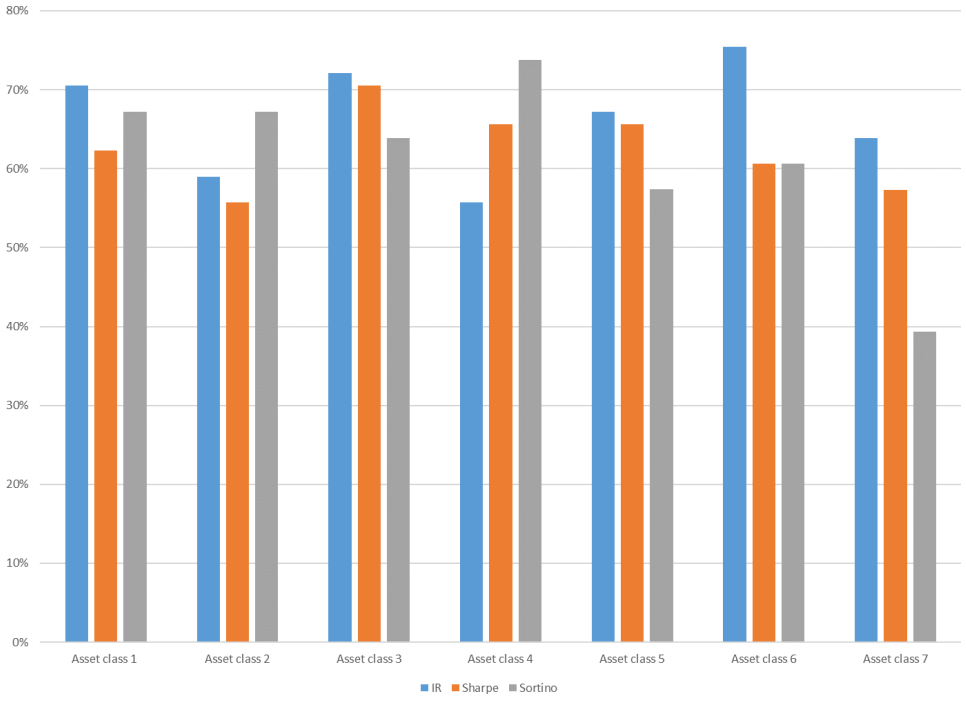

Fig. 13. Percentage of time windows in which portfolios of 4- to 5-star cardinal funds outperform the portfolios of 4- to 5-star ordinal funds over the three-year period following the fund attribution.

portfolio was subject to 64 re-compositions. Empirical evidence confirms the superiority of the cardinal approach.

Figure 14 shows, for each asset class, the average (blue line) differential return recorded in the year following the setup, between the portfolios consistent with the cardinal criterion and those consistent with the ordinal approach; differential returns are all positive, with the sole exception of Japan equities. The aforementioned figure also shows, for each asset class, the average differential return calculated considering only the time windows in which the portfolio selected via the cardinal criterion beats those consistent with the ordinal approach (green line) and vice versa (red line).

The third empirical test was conducted using the *contingency table* technique widely used to measure the reliability of credit ratings in anticipating the default of companies. In this case, the *default* event corresponds to the manager's inability to beat the reference market. This methodology is already widely used in the field of asset management for the purpose of verifying the performance persistence of asset managers (Sharpe 1966, Goetzmann & Ibbotson 1994). In this work, as already specified, the classification between *winner* and *loser* funds is based on the attribution of a rating that is assumed to be good (≥ 4 stars) or bad (≤ 3 stars). In order to measure the frequency with which the funds belonging to the two groups (*best* and *worst*) experienced the *default* event (i.e. an underperformance relative to the

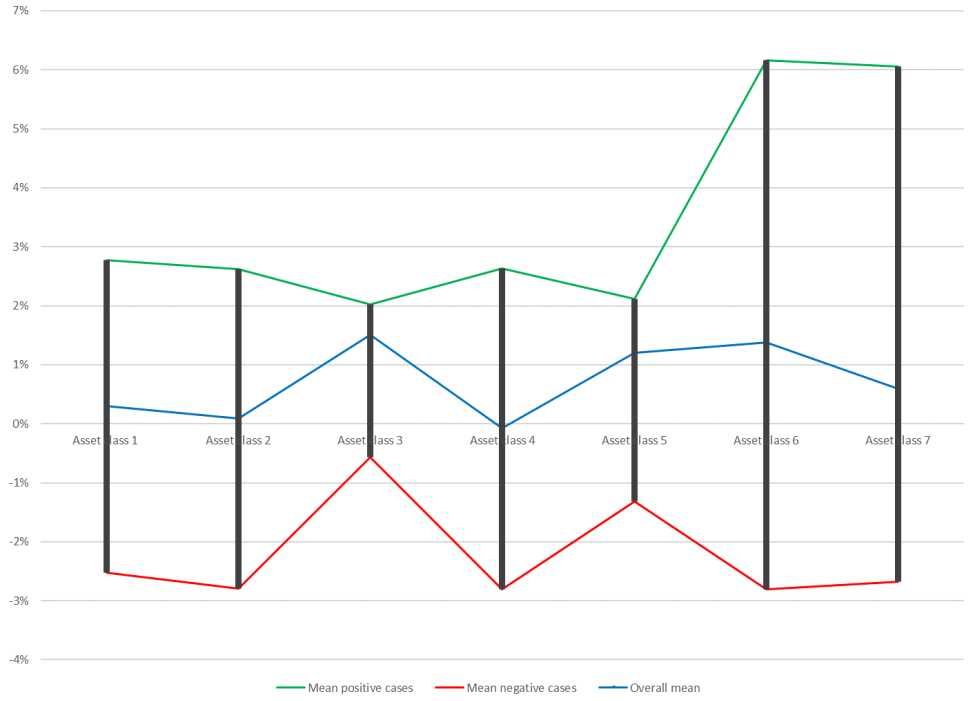

Fig. 14. Average of annual rolling observations of the differential performance between the cardinal-based portfolios and the ordinal-based ones.

market) over the entire time window (30 March 2001–2 June 2022), we attributed the ratings based on the performance of a three-year time window; then, a following 3-year time window is used to test the ability of funds to outperform the benchmark. This analysis was carried out for all nonoverlapping three-year windows of the time series. As in the previous test, the rating calculation was updated using a 13-week rolling window. Values thus obtained were expressed in percentage terms so as to allow a comparison even in the presence of a different number of *best* and *worst* funds. A "negative surprise" is the event in which a *best* fund did not outperform the market in the three-year period following the evaluation period, and a "positive surprise" is the event in which a *worst*fund outperformed the market. The two other scenarios (*best* and *worst* funds, respectively, outperforming and underperforming the benchmark) are confirmations of the assessment made in advance. Table 6 shows the structure of the contingency table and summarizes all the measures used to quantify the reliability of rating methods. Table 7 shows the results for all asset classes.

For Europe equity, we observed that 51.2% of funds with 4 or higher cardinal stars beat the market; this percentage drops to 43.3% for the ordinal criterion. Similarly, the percentage of less than 4 cardinal stars funds underperforming the market is 68.6%; for the ordinal approach, the percentage rises to 73%. The Accuracy Rate (best funds outperforming or worse funds underperforming the market) is

Table 6. Contingency table and measures.

Predicted Rank	Past Performance: Positive (IR>0)	Past Performance: Negative (IR<0)	Measures:	Measures:
Positive (rating>3)	Positive-Positive (PP)	Negative Surprise (NS)	*Positive Prediction Ratio (* PPR*):* PP/ (PP+NS)	*(1-* PPR*):* NS/ (PP+NS)
Negative (rating<=3)	Positive Surprise (PS)	Negative-Negative (NN)	*Negative Prediction Ratio (* NPR*):* NN/ (PS+NN)	*(1-* NPR*):* PS/ (PS+NN)
Measures:	*Sensitivity* : PP/ (PP+PS)	*Specificity* : NN/ (NS+NN)	*Accuracy* : (PP+NN)/ (PP+NS+PS+NN)	

66.7% with the cardinal criterion and 61.1% with the ordinal one. It is meaningful to highlight that it is more important to focus on negative surprises rather than on positive ones. Since the rating is used to select the best funds, the "real" risk the investor runs is the selection of a positively rated fund that then performs badly; on the other hand, a negatively rated fund that performs well has no real impact on the investor's performance. Focusing on the most significant error, the gap between the two criteria widens: the percentage of negative surprises ($= 1 - \text{PPR}$) is 48.8% for the cardinal criterion compared to 56.7% for the ordinal criterion.

In the case of the North American equity, both criteria show a high number of surprises; a particularly high percentage of negative surprises is observed: 74.5% with the cardinal criterion and 81.7% with the ordinal one. This result is consistent with the already mentioned higher information efficiency of the North American stock market, which explains the lower percentage of funds able to outperform the market. Again, the Accurate Rate rewards the cardinal criterion (83.1% versus 61.0%).

In the case of Japan and Emerging Markets Equity, minor differences are found between the two rating criteria. However, even in these cases, the cardinal approach is able to minimize negative surprises (46.8%/64.7% versus 49.4%/68.5%), thus confirming a better predictive capacity to select the "best" funds. Cardinal Accuracy Ratios (65.7%/71.3%) are higher than the Ordinal ones (64.9%/60.2%).

As far as the bond asset classes are concerned, *contingency table* shows homogeneous results: due to a lower incidence of negative surprises, total surprises with the cardinal criterion are systematically lower than the ones with the ordinal criterion.

Table 7. Contingency measures for all asset classes.

Category/Asset Class	Num. of cases	Rating method	Sensitivity	Specificity	Accuracy	PPR	NPR	1-PPR	1-NPR
Europe equity	7,562	Cardinal:	16.7%	92.0%	66.7%	51.2%	68.6%	48.8%	31.4%
		Ordinal:	51.6%	65.9%	61.1%	43.3%	73.0%	56.7%	27.0%
North America equity	5,007	Cardinal:	12.2%	94.3%	83.1%	25.5%	87.1%	74.5%	12.9%
		Ordinal:	53.4%	62.2%	61.0%	18.3%	89.4%	81.7%	10.6%
Japan equity	3,292	Cardinal:	30.1%	85.3%	65.7%	53.2%	68.8%	46.8%	31.2%
		Ordinal:	56.8%	69.4%	64.9%	50.6%	74.4%	49.4%	25.6%
Emerging Markets equity	5,047	Cardinal:	17.2%	89.4%	71.3%	35.3%	76.4%	64.7%	23.6%
		Ordinal:	50.4%	63.4%	60.2%	31.5%	79.3%	68.5%	20.7%
Euro bond	5,903	Cardinal:	15.0%	92.6%	77.9%	32.2%	82.4%	67.8%	17.6%
		Ordinal:	56.1%	63.8%	62.3%	26.6%	86.2%	73.4%	13.8%
Global bond	6,329	Cardinal:	20.9%	88.9%	61.0%	56.7%	61.8%	43.3%	38.2%
		Ordinal:	49.6%	66.5%	59.6%	50.7%	65.5%	49.3%	34.5%
Emerging Markets bond	3,709	Cardinal:	17.6%	90.8%	74.2%	36.0%	79.0%	64.0%	21.0%
		Ordinal:	50.7%	63.1%	60.3%	28.7%	81.4%	71.3%	18.6%
	Arithmetic	Cardinal:	18.5%	90.5%	71.4%	41.4%	74.9%	58.6%	25.1%
	Mean	Ordinal:	52.6%	64.9%	61.3%	35.7%	78.4%	64.3%	21.6%

Table 8. The market neutral home bias adjusted asset allocation.

Category/Asset class	Weight (%)
(1) Europe equity	15
(2) North America equity	25
(3) Japan equity	6
(4) Emerging Markets equity	4
(5) Euro bond	31
(6) Global bond	15
(7) Emerging Markets bond	4

The aggregate analysis (arithmetic mean) provided by *contingency table* ultimately leads to the following remarks:

– the cardinal criterion proves to be better as the percentage of surprises is significantly lower;
– both criteria have a low predictive capacity and do not guarantee at all that the funds labeled as best will beat the market;
– the cardinal criterion seems to be more effective in minimizing negative surprises, the only ones that really affect the quality of the investor's portfolio.

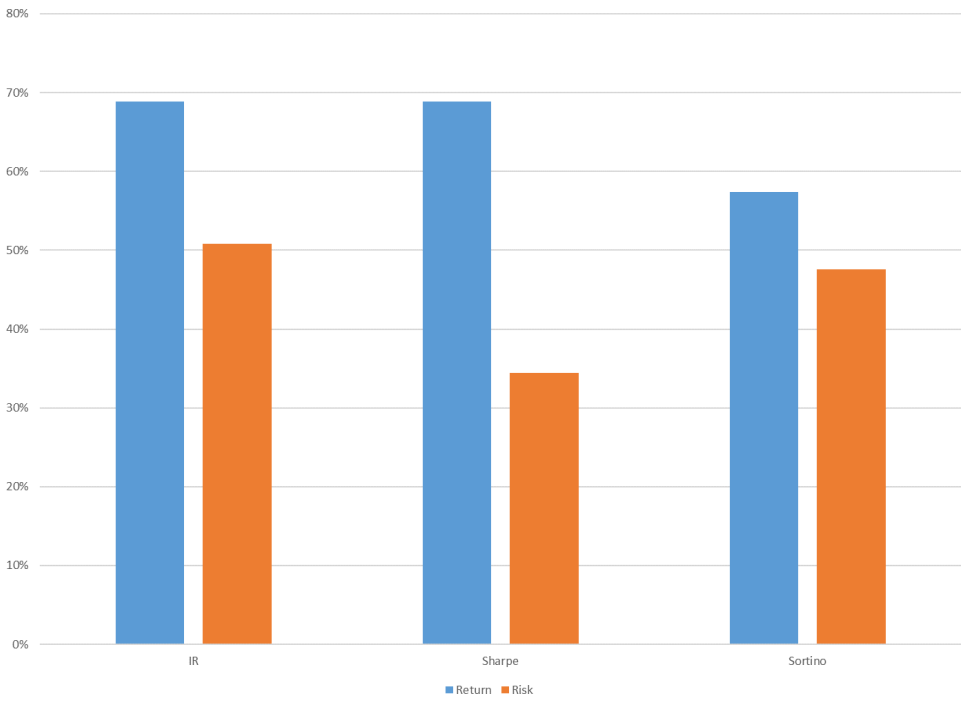

Fig. 15. Percentage of cases in which the cardinal-based portfolio outperforms the ordinal-based portfolio.

7. A Portfolio Approach

The analyses in the previous paragraphs have been based on single-asset classes. Here, we perform an empirical analysis based on a portfolio approach which, in line with the literature (Brinson *et al.* 1991) that identifies strategic asset allocation as the step best able to explain the performance of an investment, aims at assessing the impact produced by a selection of multi-sector managers characterized by the highest ratings. Consistent with the Capital Asset Pricing Model, a *market neutral* portfolio was constructed, which was then adjusted to take into account the *home bias*, by giving greater weight to Euro bonds and European equities (Table 8).

The strategic asset allocation thus defined was then implemented by choosing the funds belonging to the peer groups in accordance with a constant-mix approach that assumes a periodic rebalancing of the portfolio. Within each peer group, funds characterized by a rating higher than 3 stars were equally weighted.

The first empirical test was conducted using a three-year rolling period over the entire time series (30 March 2001–2 June 2022) with the aim of quantifying the frequency with which the cardinal criterion beats the ordinal criterion in terms of return. The results obtained (Fig. 15) are in line with those observed for single asset classes (Fig. 13), confirming a superiority of the cardinal criterion based on the *information ratio:* in 68.8% of cases, a fund portfolio based on the cardinal approach beat the portfolio consistent with the ordinal criterion. The superiority of the cardinal criterion is also confirmed when using the Sharpe Index and Sortino Index. The comparison in terms of risk, with the only exception of the standard deviation, shows a substantial parity (values close to 50%) between the criteria.

One more empirical test was conducted at portfolio level, to verify whether the cardinal-based portfolios were able to beat those based on the ordinal criterion in the year following the valuation date. Based on the 15 rolling periods, the results obtained reward the cardinal criterion, which records not only nine outperforming periods (with an average value of positive differential returns equal to +2.16%) but also an average differential return by 0.59%.

8. Conclusions

The mutual fund rating is at the center of the scientific debate, and the academic literature has proposed numerous risk-adjusted return measures to analyze topics like the differential return (with respect to the benchmark or the risk-free rate) and the risk measurement. Less investigated is the issue of how these indicators can be used to attribute a synthetic measure (rating) that allows an immediate and intuitive assessment of a manager's performance. This work aims to question the widespread recourse to the ordinal criterion of rating assessment, which is implicitly based on the principle, not fully shareable, that a fund can receive a positive rating (equal to or higher than 3) regardless of both its absolute return and its ability to beat the reference market. The application of this criterion carries the risk of being labeled as

excellent funds that, although performing better than competitors, were not able to show a positive tracking error. In our study, we have, therefore, analyzed how the rating attribution system changes by introducing not only a cardinal classification, but also a filter that allows a positive rating (equal to or greater than 3) to be attributed only to those funds that beat the market, if active, or replicate it with below-average cost, if passive.

The empirical tests we conducted allowed us to assess the impact of this new approach both *in-sample* and *out-of-sample*. We first analyzed the three-year period from 31 August 2019 to 2 June 2022 for funds investing in seven asset classes, providing a broad coverage of investment opportunities in the equity and bond markets. Funds were rated with cardinal and ordinal criteria using three measures of *risk-adjusted performance*.

Empirical evidence highlighted the risk of attributing positive ratings to funds performing worse than the market due to poor investment choices and/or excessive costs. A first advantage of the cardinal criterion is its greater selectivity, limiting the number of funds labeled with a positive rating.

Interesting evidences also emerge from the *out-of-sample* analysis conducted by broadening the time window. Covering a period of more than 20 years (from 30 March 2001 to 2 June 2022), we performed a *rolling* analysis aimed at verifying whether the ratings assigned using the two criteria (cardinal *versus* ordinal) were able to anticipate the future performance of the funds. The first analysis concerned the performance, in the quarter following the rating attribution, of portfolios of 4- and 5-star funds set up on the basis of the cardinal and the ordinal criteria. The comparison favors the cardinal approach, particularly for asset classes with higher levels of information efficiency.

Subsequently, we extended the time horizon of the *out-of-sample* analysis by comparing the return achieved by portfolios constructed with the cardinal/ordinal criterion over the 12 months following the rating assignment. Here too, it emerges that the cardinal criterion yields on average better results.

A further empirical test has been conducted via the construction of *contingency tables* identifying the default event in the inability to beat the benchmark and thus allowing an assessment of the confirmations and errors made by the rating system. In particular, the analysis focuses on negative surprises, i.e. the circumstance that funds rated positively (rating ≥ 3) are in the following three years unable to retain the same rating. Evidences confirm the research question: the cardinal criterion makes it possible to reduce the percentage of "negative surprises".

Finally, we switched from a single asset class to a market neutral home bias adjusted asset allocation. Empirical evidences confirm the superiority of the cardinal criterion also at the portfolio level.

We believe that both companies assigning ratings and financial institutions identifying the *best in class* fund should take into consideration, in addition to the *risk-adjusted performance* measure, the rule used to assign the rating, since our analysis shows that switching the criterion has significant consequences in rating

assignment. The application of a cardinal criterion, capable of significantly limiting the subset of funds labeled as high-ranked, should help to enhance the future performance of portfolios.

References

G. P. Brinson, B. D. Singer & G. L. Beebower (1991) Determinants of portfolio performance II: An update, *The Financial Analyst's Journal* **47** (3), 40–48.

K. C. Brown, W. V. Harlow & L. T. Starks (1996) Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry, *The Journal of Finance* **51** (1), 85–110.

D. Del Guercio & P. A. Tkac (2008) The effect of morningstar ratings on mutual fund flow, *The Journal of Financial and Quantitative Analysis* **43** (4), 907–936.

Esma (2018).

W. N. Goetzmann & R. G. Ibbotson (1994) Do winners repeat? *The Journal of Portfolio Management* **20** (2), 9–18.

http://www.lipperleaders.com/documents/lipperleaders_methodology_us_10mar09.pdf.

https://www.morningstar.com/content/dam/marketing/shared/research/methodology/771945_Morningstar_Rating_for_Funds_Methodology.pdf.

https://www.quantalys.it/Whoarewe/Methodology.

H. M. Markowitz (1952) Portfolio selection, *The Journal of Finance* **VII** (1), 77–91.

H. M. Markowitz (1959) *Portfolio Selection. Efficient Diversification of Investments.* New York: John Wiley & Sons.

F. Modigliani & L. Modigliani (1997) Risk adjusted performance, *The Journal of Portfolio Management* **23** (2), 45–54.

A. Peiro (1994) The distribution of stock returns: International evidence, *Journal of Applied Financial Economics* **4** (6), 431–439.

A. Peiro (1999) Skewness in financial returns, *Journal of Banking and Finance* **23** (6), 847–862.

W. F. Sharpe (1964) Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* **19** (3), 425–442.

W. F. Sharpe (1966) Mutual fund performance, *Journal of Business* **39** (1), 119–138.

W. F. Sharpe (1991) The arithmetic of active management, *The Financial Analyst's Journal* **47** (1), 7–9.

W. F. Sharpe (1994) The Sharpe ratio, *The Journal of Portfolio Management* **21** (1), 49–58.

J. Singleton & J. Wingender (1986) Skewness persistence in common stock returns, *Journal of Financial and Quantitative Analysis* **21** (3), 335–341.

F. A. Sortino & S. Satchell (2001) *Managing Downside Risk.* Oxford: Butterworth-Heinemann.

F. A. Sortino & R. van der Meer (1991) Downside risk, *The Journal of Portfolio Management Summer* **17** (4), 27–31.

L. T. Starks (1987) Performance incentive fees: an agency theoretic approach, *Journal of Financial and Quantitative Analysis* **22** (1), 17–32.

C. Tobe (1999) Tracking error: An essential tool in evaluating index funds, *Journal of Investment Consulting* **2** (1), 8–11.

J. L. Treynor (1965) How to rate management of investment funds, *Harvard Business Review* **43** (1), 63–75.

T. W. Young (1991) Calmar ratio: A smoother tool, *Futures* **20** (1), 40–41.