INTERNATIONAL JOURNAL OF FINANCE AND ECONOMICS
Int. J. Fin. Econ. **17**: 61–72 (2012)
Published online 12 December 2010 in Wiley Online Library
(wileyonlinelibrary.com). DOI: 10.1002/ijfe.437

YES, THE CHOICE OF PERFORMANCE MEASURE DOES MATTER FOR RANKING OF US MUTUAL FUNDS[‡]

JOSÉ RENATO HAAS ORNELAS[a,*,†], ANTÔNIO FRANCISCO SILVA JÚNIOR[a,b,§,¶] and JOSÉ LUIZ BARROS FERNANDES[a]
[a]*Banco Central do Brasil, Brazil*
[b]*Universidade Federal da Bahia, Brazil*

ABSTRACT

Recent literature in performance evaluation has focused on preferences and characteristics of returns' distribution that go beyond mean and variance world. However, Eling (2008) compared the Sharpe ratio with some of these performance measures, and found virtually identical rank ordering using mutual fund data. This paper compares 13 performance measures with the traditional Sharpe Ratio using a sample of US Fixed-Income, Equity and Asset Allocation Mutual Funds. Results show that performance measures based on absolute reward-risk ratios have similar rankings, when the numerator (mean excess return) is the same. However, when we move to other types of performances measures, results may be significantly different. This is the case of the Manipulation-Proof Performance Measure (MPPM), Upside Potential Ratio, and Appraisal Ratio. Results are especially different for the MPPM. Robustness checks show that some of the performance measures are very sensitive to parameters' changes. Therefore, the choice of the performance measure is actually important for mutual fund ranking and selection. As a consequence, we argue that the use of several performance measures and rankings have a positive impact on the mutual fund's industry, reducing concentration. Copyright © 2011 John Wiley & Sons, Ltd.

JEL CODE: D81; G11

KEY WORDS: Performance measure; rank correlation; sharpe ratio

1. INTRODUCTION

The Sharpe Ratio is the most widely known and used performance measure for the Mutual Fund Industry. It measures the relationship between the mean excess return (risk premium over the default-free short-term interest rate) and the standard deviation of the returns generated by the fund (Sharpe, 1966).

But the Sharpe Ratio is an adequate measure of performance evaluation only when investors believe that risk can be properly measured by standard deviation, or in a world where returns have nice symmetric distributions like the normal. In the real world, we can find several categories of funds with non-Normal shapes.

One straightforward reason for these odd-shaped distributions of funds is that individual assets available have, themselves, odd-shaped distributions. Thus, when these assets are included in fund's portfolio, the resulting distribution is non-Normal.

Another possible reason for odd-shaped distributions in funds is the use of the so-called 'Portfolio Manipulation'. Fund managers change frequently the portfolio composition and leverage and this leads to a distribution variation across time, originating an odd-shaped unconditional distribution. As Sharpe Ratio

*Correspondence to: José Renato Haas Ornelas, Banco Central do Brasil, Brazil.
†E-mail: jrenato.ornelas@bcb.gov.br
‡The views expressed in this work are those of the authors and do not reflect those of the Banco Central do Brasil or its members.
§Correspondence to: Antônio Francisco Silva Júnior, Banco Central do Brasil, Brazil.
¶E-mail: antonio.francisco@bcb.gov.br

and many other measures assume stationary distributions, ranking results of such measures may be misleading. Goetzmann *et al.* (2007) call 'Dynamic Manipulation' the strategies that induce these temporal effects in order to obtain higher measures of performance.

Fund managers can also generate odd-shaped distributions using what Goetzmann *et al.* (2007) call 'Static Manipulation'. These strategies generate several small excess returns and few big losses, creating a distribution with negative skewness. For instance, a simple strategy of selling out-of-the-money calls during several expiration cycles may generate high Sharpe Ratios.

There is a growing literature of performance measurement that goes beyond the mean-variance world, and in some sense try to overcome Sharpe Ratio's manipulation problems. These numerous performance measures include Sortino Ratio (Sortino and Van der Meer, 1991), the Calmar Ratio (Young, 1991), Omega ratio (Keating and Shadwick, 2002), the Manipulation-Proof Performance Measure (MPPM) (Goetzmann *et al.*, 2007), and the measures proposed by Koekebakker and Zakamouline (2009). These measures try to capture information about the fund that goes beyond the mean return and standard deviation of returns.

However, Eling (2008) compared the Sharpe ratio with some of these performance indexes, and found virtually identical rank ordering by these measures, using a sample of mutual funds across the world. Although this paper has compared 10 performance measures with the Sharpe Ratio, they did not consider some important measures. For instance, they did not consider a measure that has been proposed in parallel by both academics and practitioners: the MPPM or MorningStar Risk-Adjusted Performance Measure (MSRAPM, see Morningstar, 2006). This performance measure goes beyond the mean-variance world and according to the authors is not prone to manipulation strategies commonly employed to game Sharpe Ratio and related measures.

For Hedge Fund data, the paper of Eling and Schuhmacher (2007) uses several performance measures and concludes that rankings are virtually identical. On the other hand, Zakamouline (2009) uses the same Hedge Fund data, and find that for some alternative performance measures ranking may be different.

In order to shed more light in the previous contradiction, our paper evaluates empirically the ranking correlation of the Sharpe Ratio with 13 performance measures, including MPPM proposed by Goetzmann *et al.* (2007). We use a sample of US mutual fund data. The objective is to check whether the use of these measures alters significantly the ranking of mutual funds. Our results indicate that the choice of performance measure is important to the ranking of funds, and thus is in line with Zakamouline (2009), but contrast with those of Eling and Schuhmacher (2007) and Eling (2008).

The remaining of the paper is organized as follow: Section 2 reviews the literature of Performance Measures, addressing both classic and new measures; Section 3 presents a summary of returns' data; Section 4 compares rankings generated by Sharpe Ratio and other measures; Section 5 makes some robustness checks and finally we present our conclusions in Section 6.

2. LITERATURE REVIEW OF PERFORMANCE MEASURES

Performance evaluation is a controversial task. We can find several measures to rank mutual and hedge funds. Cogneau and Hübner (2009a,b) describe more than 100 performance measures. It is worth to say that although ranking managers give information for investors' decision, the differences between managers may not be captured by a single measure in a short time window. According to his literature survey, Géhin (2004) shows some factors that influence performance evaluation such as data quality, survivorship bias, instant history bias, funds' size, funds' age, market factors, returns probability distribution, etc. In fact, we argue that rankings may also depend on the kind of measure we use.

Indeed, Sharpe Ratio is probably the most widely used measure of financial performance. However, because it is based on mean-variance theory, it is a meaningful measure of performance when either risk perceived by investors can be expressed exclusively by standard deviation or when returns are normally distributed. Other measures, such as the Treynor Ratio, are also based on the mean–variance world, although they also focus on other aspects of performance. Recently, there is a growing literature on performance evaluation that tries to take into account higher moments of distribution. We see two reasons

DOI: 10.1002/ijfe

for the emergence of these measures: first, there is a new paradigm of investors' perception of risk that goes beyond the variance; and second, many asset return's distributions have actually non-normal distributions.

The first reason is linked to the increasing use in the last 15 years of a number of risk measures that focus on the left tail of return's distribution, like the Value at Risk (VaR), Expected Shortfall and others. This leads to a search for performance measures that consider these kinds of risk measures. Sharpe Ratio is commonly interpreted as a reward-to-risk ratio. Thus, many researches replace standard deviation in the Sharpe ratio by risk measures that focus on the left tail of distributions. Sortino and Van der Meer (1991) replace the standard deviation by the downside deviation. Dowd (2000) uses the Value-at-Risk (VaR) measure instead of standard deviation. Stutzer (2000) proposes the Stutzer index, which is based on the assumption that investors want to minimize the probability of underperforming a specific benchmark. Keating and Shadwick (2002) introduced the Omega Ratio that is defined as the ratio of the gain relative to a given threshold to the loss with respect to the same threshold, considering the distribution of fund's returns. As pointed out by Koekebakker and Zakamouline (2008), ranking of portfolios based on this measure depends heavily on the choice of a threshold.

The second reason why performance measures are going beyond mean-variance is because assets, portfolios and funds return's distributions are actually not normally distributed. Hedge funds are an example of odd-shaped distributions that deviate substantially from normality (see, for example Malkiel and Saha, 2005). Nevertheless, Sharpe Ratio and similar measures are commonly used to evaluate and rank hedge funds.

Returns' distribution of a fund may be non-normal simply because fund's holdings contain assets with non-normal properties. However, several papers have argued that hedge fund managers use strategies trying to manipulate Sharpe-like performance measures. Goetzmann *et al.* (2007) describe three general strategies for manipulating a performance measure. The first strategy is the manipulation of the underlying distribution in order to influence the measure. The second one is dynamic manipulation that induces time variation into the return distribution to influence measures that assume stationarity. The last strategy is dynamic manipulation that focuses on inducing estimation error.

In fact, major papers in the recent performance measurement literature try to capture higher moments of returns' distributions. Many approaches are derived from utility functions that take into account investors' preference to higher moments of returns distributions. Hodges (1998) proposes an approach that generalizes the Sharpe Ratio for all moments of the distribution. However, his measure can be computed only numerically. Koekebakker and Zakamouline (2009) propose measures that take into account skewness and kurtosis preferences in investment decisions. They argue that investors are not neutral to skewness and kurtosis and so there is a premium associated to negative skewness and positive excess kurtosis. One of their measures incorporates Skewness to the Sharpe Ratio and has a closed-form formula. Nevertheless, it has a major drawback: it provides meaningful numbers only when skewness and Sharpe Ratio are inside a specific range. In some cases this measure generates imaginary numbers, and this is the case of the dataset used in our paper.

Goetzmann *et al.* (2007) present numerical simulations of manipulation strategies for several performance measures such as Sharpe Ratio, Treynor Ratio, Appraisal Ratio, Sortino and Van der Meer (1991), Sortino, Van der Meer and Plantinga (1999) ratio, and the timing measures of Henriksson and Merton (1981), and Treynor and Mazuy (1966). Manipulations presented generate statistically and economically significant better portfolio scores even though an uninformed investor may have performance equal to that obtained by all the simulated trades. Goetzmann *et al.* (2007) suggest that MPPM should have four properties: (i) produce a single valued score with which to rank each subject; (ii) score's value should not depend on portfolio's size; (iii) an uninformed investor cannot expect to enhance his estimated score by deviating from the benchmark and at the same time informed investors should be able to produce higher scoring portfolios by using arbitrage; and (iv) measure should be consistent with standard financial market equilibrium conditions. Therefore, they present MPPM.

Despite this wide variety of measures, Eling (2008) analyze and compare 11 different performance measures for a data set of mutual fund returns and conclude that all these performance measures produce virtually identical rankings. He used the following measures: Sharpe Ratio (Sharpe, 1966), Omega Ratio (Shadwick and Keating, 2002), Sortino Ratio (Sortino and Van der Meer, 1991), Upside Potential Ratio (Sortino, Van der Meer and Plantinga, 1999), Kappa 3 (Kaplan and Knowles, 2004), the Calmar

DOI: 10.1002/ijfe

Ratio (Young, 1991), the Sterling Ratio (Kestner, 1996), the Burke Ratio (Burke, 1994), the excess return on Value at Risk (Dowd, 2000), the conditional Sharpe Ratio (Agarwal and Naik, 2004), and the modified Sharpe Ratio (Gregoriou and Gueyie, 2003).

Besides ten of the eleven measures used in the Eling (2008) article, we evaluated in our paper four other measures: the MPPM, Treynor index (Treynor, 1965), Appraisal Ratio (Treynor and Black, 1973), and the Generalized Sharpe Ratio (Sharpe, 1994). Although the MPPM is a relative new measure, accordingly Goetzmann *et al*. (2007) it is similar to the MSRAPM that is also used by practitioners. The Generalized Sharpe Ratio calculates a mean and variance ratio using differential returns, i.e. returns relative to a benchmark index. In fact, our goal is to discuss the question raised by Eling (2008): 'Does the Measure Matter in the Mutual Fund Industry?'. In order to discuss this question, we evaluate the rankings generated by the following performance measures:

Performance measure	Formula
Sharpe Ratio	$(r_P - r_f)/\sigma_P$
Omega Ratio	$(r_P - \tau)/\mathrm{LPM}_{Pn}(\tau)+1$
Sortino Ratio	$(r_P - \tau)/\sqrt{\mathrm{LPM}_{P2}(\tau)}$
Upside Potential Ratio	$HPM_{P1}(\tau)/\sqrt{\mathrm{LPM}_{P2}(\tau)}$
Kappa3	$(r_P - \tau)/\sqrt[3]{\mathrm{LPM}_{P3}(\tau)}$
Calmar Ratio	$(r_P - r_f)/(-D_{P1})$
Sterling Ratio	$(r_P - r_f)/\left[(1/K)\sum_{k=1}^{K} -D_{Pk}\right]$
Burke Ratio	$(r_P - r_f)/\sqrt{\sum_{k=1}^{K} D_{Pk}^2}$
Dowd	$(r_P - r_f)/\mathrm{VaR}$
Conditional Sharpe Ratio	$(r_P - r_f)/\mathrm{CvaR}$
Generalized Sharpe Ratio	$(r_P - r_B)/\sigma_{r_P - r_B}$
MPPM	$\dfrac{1}{(1-\rho)\Delta t}\ln\left(\dfrac{1}{T}\sum_{t=1}^{T}[(1+r_t)/(1+r_{ft})]^{(1-\rho)}\right)$
Treynor Index	$r_P - r_f/\beta_P$
Appraisal Ratio	$\alpha_P/\sigma(\varepsilon_P)$

where r_P is the average compound return, r_f is the risk-free interest rate, σ_P the standard deviation of portfolio returns, LPM_{Pn} the lower partial moment of order n, equal to $(1/T)\sum_{t=1}^{T}\max(\tau - r_{Pt}, 0)^n$, with τ as the minimum acceptable return, HPM_{Pn} the higher partial moment of order n, equal to $(1/T)\sum_{t=1}^{T}\max(r_{Pt} - \tau, 0)^n$, D_k the drawdown of the portfolio returns, k the number of drawdowns (one is the maximum drawdown, is the second-largest drawdown, and so on), VaR the value at risk, calculated using a Normal distribution with historical standard deviation, CVaR conditional value at risk $(E(-r_{Pt}|r_{Pt}\leqslant -\mathrm{VaR}))$, β_P the portfolio's beta, ε the residuals of regression, $r_{P,i}$ portfolio return in date i, $r_{B,i}$ the benchmark return in date i, T the total number of observations, Δt the length of time between observations, and ρ is a constant representing the relative risk aversion.[1]

3. DATA

Data from mutual funds was obtained from Thomson Datastream and consist of monthly and daily total return indexes of US Fixed-Income, Equity and Asset Allocation Mutual funds from January 1998 to August 2008. Classification into Equity, Fixed-Income, and Asset Allocation was done using information from Bloomberg, more specifically the 'Asset Class Focus' classification from the FSRC function. We have collected

DOI: 10.1002/ijfe

Table 1. Main characteristics of the sample January 1998 to August 2008

Time-series properties by fund type	Cross-section properties			
	Mean	Standard Deviation	Minimum	Maximum
Asset allocation (273 funds)				
Mean	0.125%	0.223%	−0.800%	1.292%
Standard Deviation	3.352%	1.014%	0.988%	8.680%
Skewness	−0.72	0.88	−8.53	5.27
Kurtosis	3.013	7.00	−0.59	87.44
Equity (2061 funds)				
Mean	0.118%	0.403%	−2.801%	8.425%
Standard Deviation	5.705%	2.304%	1.210%	72.223%
Skewness	−0.46	0.89	−8.95	10.98
Kurtosis	2.811	6.91	−0.70	123.58
Fixed-Income (1521 funds)				
Mean	0.323%	0.147%	−0.619%	2.967%
Standard Deviation	1.386%	2.445%	0.238%	68.936%
Skewness	−0.66	1.09	−10.11	11.40
Kurtosis	3.643	9.81	−0.68	129.97

this information as a snapshot at the end of the sample period, and this may be a source of error. We have also collected in Bloomberg information about the Fund's primary Benchmark (FUND_BENCHMARK_PRIM field). Fund of funds and Index Funds were excluded to avoid spurious clustering. The sample included only surviving funds, i.e. funds with data for the full period. We excluded from the sample funds that do not have information for more than 1% of the business days. The initial list of Datastream has 21 246 mutual funds. We have available 3875 mutual funds. Most of the funds have been excluded because they started operations after 1998. The risk-free asset used was the 1–3 months T-Bill index of Barclays Capital.

Given these exclusions, our sample of funds has a survivorship bias, i.e. funds that disappeared before the end of the sample were not included. But as our concern in the study is to evaluate the ranking changes depending on the performance measure, this is not a relevant problem. In fact, the only way to compare the performance of funds through a specific time interval is to use only funds alive in the whole period. However, one has to be careful in analyzing the average performance of funds against a benchmark. As funds tend to be closed after negative results, the surviving funds will have an upward bias in their performance compared with the full sample.

Table 1 shows the main characteristics of the Samples. We have calculated the compound mean return, standard deviation, skewness, and kurtosis for each fund's monthly return's time series. Then, for each of these characteristics and grouped by type of fund, we calculated the cross section mean, standard deviation, minimum, and maximum values.

From Table 1 we can observe that Equity funds present, as usual, a higher volatility than the remaining ones. However, the risk premium was not observed as the Equity funds' return was not higher than the others. A possible reason for this fact is the period considered, which ended in the international financial market crises of 2008 that penalized deeply equity returns in favor of the bonds market (flight to quality effect). Related to skewness and kurtosis, we can observe that, in general, all samples present non-normal returns with negative skewness and excess kurtosis, highlighting the need for performance measures that go beyond the Sharpe Ratio.

4. RANKING COMPARISON

This section compares funds' performance using 13 different performance measures described in Section 2 with the Traditional Sharpe Ratio. We do it using a monthly sample starting in 1998. In order to compare the measures we first rank all funds using the measures, grouped by type of fund and also grouped by the

DOI: 10.1002/ijfe

Table 2. Spearman's rank correlation using monthly data

Performance measure	Equity	Fixed-income	Asset allocation
(A) Spearman's rank correlation—grouped by type			
MPPM	63.18%	95.83%	81.10%
Omega	99.84%	99.58%	99.63%
Sortino	99.92%	99.64%	99.84%
Kappa	99.58%	99.54%	98.72%
Upside Potential Ratio	90.51%	89.43%	81.75%
Calmar	98.16%	98.73%	95.38%
Sterling	99.58%	99.30%	98.52%
Burke	99.45%	99.24%	98.27%
Dowd	99.46%	98.87%	98.73%
Conditional Sharpe Ratio	99.69%	99.42%	98.93%

Performance measure	S&P500	MSCI EAFE	Russell 2000	Russell 1000 Value	Russell 1000 Growth
(B) Spearman's rank correlation—grouped by benchmark					
MPPM	74.17%	82.04%	73.34%	75.76%	87.15%
Omega	99.85%	99.93%	99.82%	99.24%	99.34%
Sortino	99.91%	99.95%	99.92%	99.77%	99.80%
Kappa	99.46%	99.81%	99.75%	98.59%	98.82%
Upside Potential Ratio	89.95%	87.01%	83.77%	86.46%	90.54%
Calmar	97.95%	98.90%	98.96%	95.01%	90.00%
Sterling	99.52%	99.76%	99.66%	98.79%	99.15%
Burke	99.38%	99.64%	99.59%	98.75%	98.43%
Dowd	99.48%	99.81%	99.09%	98.25%	98.99%
Conditional Sharpe Ratio	99.61%	99.88%	99.73%	99.46%	98.90%
Treynor	93.50%	98.98%	99.05%	98.31%	98.18%
Appraisal	90.14%	85.99%	86.89%	81.97%	89.50%
Generalized Sharpe Ratio	94.60%	99.87%	94.93%	91.22%	93.22%

Benchmark. Then we calculate both the Spearman's Rho Ranking Correlation (Table 2) and the Kendall's Tau Ranking correlation (Table 5). In the next section, during the robustness analysis, we decided to use just the Spearman's Rho. The reason is that using Spearman's Rho makes our results comparable with many previous papers, including Eling (2008), although we believe that the Kendall's Tau Rank Correlation is more appropriate in this analysis.[2]

Panel A of Table 2 shows Rank Correlations of the Sharpe Ratio with the measures that do not require a Benchmark, grouped by type of fund. These measures are basically the same used by Eling (2008), and results are very similar except for the Upside Potential Ratio (UPR). The reason is probably that we consider the mean of compound returns instead of arithmetic mean as in Eling (2008) for the Share Ratio, Drawdown-based measures, and Lower Partial moment's measures. The numerator of the UPR—the Higher Partial Moment of order one—is similar to the arithmetic mean, but is less related to the compound mean. This can explain higher correlations of Eling (2008) for the UPR.

Still in Panel A, we see that the MPPM produces correlation rankings for Equity and Asset Allocation funds that are far from 100%, meaning that there is a relevant difference in these rankings.

Panel B of Table 2 shows Rank Correlations of the Sharpe Ratio with the measures grouped by the most common Benchmark indexes in our sample (all of them equities). Here we also consider measures that do require a Benchmark. Results show that for the measures also present on Panel A, we do not have many differences: UPR and MPPM have lower rank correlation. For the Benchmark-based measures, we have the Treynor and GSR with correlation over 90%, but for Appraisal Ratio the correlation is in the range 80–90%. Although it is a high correlation, it is substantially lower than the 98–99% levels of the non-benchmark-based measures. It is important to highlight that most of the funds in the Equity sample have S&P500 as their Benchmark.

DOI: 10.1002/ijfe

Table 3. Kendall's rank correlation using monthly data

Measure	Equity	Fixed-Income	Asset Allocation
(A) Kendall's rank correlation—grouped by type			
MPPM	46.21%	84.38%	65.46%
Omega	97.27%	97.40%	96.16%
Sortino	97.95%	97.95%	97.11%
Kappa	95.47%	96.56%	91.81%
Upside Potential Ratio	74.14%	75.05%	64.72%
Calmar	89.61%	91.85%	83.49%
Sterling	95.01%	94.59%	90.97%
Burke	94.35%	94.18%	90.15%
Dowd	94.44%	92.87%	91.83%
Conditional Sharpe Ratio	95.72%	95.63%	91.97%

Measure	S&P500	MSCI EAFE	Russell 2000	Russell 1000 Value	Russell 1000 Growth
(B) Kendall's rank correlation—grouped by benchmark					
MPPM	57.16%	64.68%	55.98%	55.31%	77.96%
Omega	97.29%	98.40%	97.54%	94.99%	96.59%
Sortino	97.87%	98.71%	98.13%	97.45%	96.94%
Kappa	94.97%	97.35%	96.66%	92.44%	80.90%
Upside Potential Ratio	74.08%	70.98%	67.41%	68.39%	63.33%
Calmar	88.91%	93.16%	92.39%	83.35%	71.32%
Sterling	94.58%	96.55%	96.13%	92.59%	97.06%
Burke	93.96%	95.97%	95.79%	92.24%	96.06%
Dowd	94.53%	97.13%	93.35%	91.16%	96.36%
Conditional Sharpe Ratio	95.21%	97.73%	96.52%	94.89%	98.12%
Treynor	85.59%	93.20%	94.14%	91.06%	95.18%
Appraisal	72.48%	68.13%	68.83%	61.39%	74.73%
Generalized Sharpe Ratio	81.52%	97.47%	81.36%	75.79%	80.19%

Table 3 shows the Kendall's Tau Correlation Coefficient for the same measures presented in Table 2. We see that Kendall's Tau Correlation Coefficients are in general lower than Spearman's Rho. The MPPM for equity funds has a correlation around 50%, which means three times more concordant than discordant pairs. The UPR and Appraisal measures have a correlation around 60–70%, which means five to six times more concordant than discordant pairs. The other measures have a high correlation.

In general, results show that for performance measures based on reward-risk ratios, rankings are similar when the numerator (mean excess return) is the same. This is in line with Eling (2008). However, when we move to other types of performances measures, like the MPPM, Upside Potential Ratio, and Appraisal Ratio, results clearly have not 'virtually identical rank ordering' as in Eling (2008).

5. ROBUSTNESS ANALYSIS

5.1 Sub-categories

This section compares funds' ranking correlation using subdivisions of the Equity and Fixed-Income macro-classifications, still using monthly data. These subdivisions were done using the Fund Objective from Bloomberg (FUND_OBJECTIVE_LONG field). For the Equity mutual funds, we have divided into two types of sub classifications: by market capitalization (Small, Medium, or Large) and by style (Value or Growth). For the fixed-income funds, we have divided into Municipal, Government and Agencies, and Corporates.

Results are in Table 4, and are very similar to those of Table 2. For the MPPM, correlations of Equity funds are higher when we divide either by capitalization or by style. However, for Fixed-Income funds, correlations are lower when we subdivide by type of bonds.

DOI: 10.1002/ijfe

Table 4. Rank correlation using subdivisions

Measure	Small Cap	Mid Cap	Large Cap	Equity Value	Equity Growth
(A) Spearman's rank correlation—equity subdivisions					
MPPM	70.12%	69.16%	80.23%	77.16%	71.12%
Omega	99.81%	99.58%	99.40%	99.82%	99.68%
Sortino	99.88%	99.76%	99.81%	99.89%	99.88%
Kappa	99.63%	99.04%	98.42%	99.27%	99.24%
Upside Potential Ratio	83.03%	91.82%	90.54%	87.93%	90.93%
Calmar	97.88%	96.29%	93.09%	97.18%	96.33%
Sterling	99.57%	99.11%	98.72%	99.47%	99.22%
Burke	99.41%	98.78%	98.32%	99.30%	98.94%
Dowd	99.02%	98.05%	98.84%	99.02%	99.05%
Conditional Sharpe Ratio	99.61%	99.40%	98.74%	99.38%	99.47%

Measure	Municipal	Govt & Agency	Corporates
(B) Spearman's rank correlation—fixed-income subdivisions			
MPPM	89.35%	92.10%	87.66%
Omega	99.86%	99.85%	98.76%
Sortino	99.93%	99.84%	98.80%
Kappa	99.89%	99.63%	98.74%
Upside Potential Ratio	90.38%	94.96%	90.21%
Calmar	99.31%	97.95%	98.07%
Sterling	99.76%	98.19%	98.43%
Burke	99.73%	98.26%	98.39%
Dowd	99.54%	97.63%	98.50%
Conditional Sharpe Ratio	99.85%	98.09%	98.65%

5.2 Changing the parameters

Another robustness check that can be done is to change the parameters used on the Drawdown, Lower Partial Moments, MPPM, and VaR measures. We are still using monthly data. Panel A of Table 5 shows results for MPPM with Relative Risk Aversion (RRA) parameter equal to 2, 3, 4, and 5. We see that as the RRA increases the correlation between MPPM and Sharpe Ratio decreases, meaning that investors more risk-averse would have very different rankings when using one or other measure.

Panel B of Table 5 shows Lower Partial Moments Performance Measures with Minimum Acceptable Return (MAR) equal to zero. The LPM of second order with MAR equal to zero is often referred as 'Semi-Standard Deviation'. For the Equity and Asset Allocation funds, the correlation has dropped between 3 and 9 percentage points. However, for the Fixed-income funds, the correlation using MAR equal to zero is very low, a little bit under 60%. Note that the MAR used for base case on Table 2 was the average T-Bill return for the period, which is one possible benchmark for Fixed-Income funds.

Panel C of Table 5 shows correlation with the number of Drawdown's for the Sterling and Burke Ratios equal to 10. Again, Results are very similar. Finally, Panel D of Table 5 shows Dowd and Conditional Sharpe Ratios with confidence levels equal to 90, 95, and 99%. Results are very similar.

We may conclude that MPPM and LPM measures may be very sensitive to parameter's changes.

5.3 Daily data

On this section, we perform a robustness check using daily data instead of monthly data, for the same time period. Panel A of Table 6 shows Spearman's Rank Correlation for three types of funds using daily data. Results are similar except for the UPR, which is lower. We believe that it has happened because we use the mean of compounded returns instead of using arithmetic mean in the Sharpe ratio as explained in Session 4. Using daily data the difference of these means is higher and correlation of the rankings is lower.

DOI: 10.1002/ijfe

Table 5. Robustness Spearman's rank correlation—grouped by type

Measure	Equity	Fixed-Income	Asset Allocation
(A) Spearman's rank correlation—MPPM			
MPPM with RRA = 2	87.36%	96.50%	86.46%
MPPM with RRA = 3	75.62%	95.83%	81.10%
MPPM with RRA = 4	63.68%	93.98%	76.80%
MPPM with RRA = 5	52.92%	92.26%	71.78%
(B) Spearman's rank correlation—LPM measures			
Omega	96.52%	59.22%	90.28%
Sortino	96.54%	59.81%	90.48%
Kappa	96.42%	59.88%	90.47%
Upside Potential Ratio	89.37%	57.27%	80.15%
Calmar	98.16%	98.73%	95.38%
(C) Spearman's rank correlation—drawdown measures			
Calmar	98.16%	98.73%	95.38%
Sterling	99.81%	99.31%	99.26%
Burke	99.67%	99.31%	98.93%
(D) Spearman's rank correlation—VaR-based measures			
Dowd @ 99%	99.20%	99.07%	96.84%
Dowd @ 95%	99.46%	98.87%	98.73%
Dowd @ 90%	99.41%	96.43%	97.96%
Conditional Sharpe Ratio @ 99%	97.15%	98.81%	92.89%
Conditional Sharpe Ratio @ 95%	99.69%	99.42%	98.93%
Conditional Sharpe Ratio @ 90%	99.87%	99.06%	99.65%

Table 6. Spearman's rank correlation using daily data

Measure	Equity	Fixed-Income	Asset Allocation	
(A) Spearman's rank correlation—daily data				
MPPM	82.82%	97.24%	83.00%	
Omega	99.72%	98.62%	99.11%	
Sortino	99.88%	99.34%	99.76%	
Kappa	97.95%	98.85%	96.17%	
Upside Potential Ratio	55.50%	50.25%	46.72%	
Calmar	96.73%	97.27%	93.63%	
Sterling	99.74%	98.82%	99.42%	
Burke	99.74%	98.85%	99.65%	
Dowd	99.36%	98.18%	98.59%	
Conditional Sharpe Ratio	99.59%	99.16%	99.70%	
Measure	S&P500	Russell 2000	Russell 1000 Value	Russell 1000 Growth
(B) Spearman's rank correlation—daily data grouped by benchmark				
MPPM	73.70%	89.40%	95.83%	71.70%
Omega	97.95%	98.97%	99.65%	91.84%
Sortino	99.16%	99.56%	99.91%	94.35%
Kappa	99.16%	99.36%	98.17%	95.08%
Upside Potential Ratio	64.78%	30.55%	81.74%	73.20%
Calmar	98.05%	98.40%	92.61%	90.92%
Sterling	98.33%	99.56%	99.48%	93.09%
Burke	99.05%	99.49%	99.48%	93.95%
Dowd	97.45%	99.07%	99.22%	90.71%
Conditional Sharpe Ratio	98.36%	99.90%	99.74%	93.40%
Treynor	94.66%	97.81%	99.39%	86.77%
Appraisal	77.10%	87.98%	41.30%	73.32%
Generalized Sharpe Ratio	81.30%	94.13%	51.39%	82.98%

DOI: 10.1002/ijfe

Panel B of Table 6 shows Spearman's Rank Correlation for four Benchmarks using daily data. Again the UPR shows lower correlation. For the Benchmark-based measure, we also have a lower correlation, sometimes near 50%. For the other measures, results are in general similar to the base sample.

6. SHOULD WE RELY ON A SINGLE PERFORMANCE MEASURE?

We can analyze results dividing performance measures in three categories. The first category comprises Omega, Sortino, Kappa, Sterling, Burke, Dowd, and Conditional Sharpe Ratio. Note that correlations with Sharpe ratios are very high for these measures. Since the numerators of these indexes are the same, we may conclude that the risk adjustment used by each measure in the denominator does not change the final rankings. It may indicate that for ranking purposes these measures lead to the same conclusions. In other words, the different risk measures do not change relative evaluation of the funds. Our findings here are similar to those of Eling (2008) and we agree with the author when he states that measures are correlated since numerator is excess return for 10 of the 11 measures used by him and the denominator contains a more or less comparable risk measure. It is important to consider the findings of Eling: *I also found high rank correlations when comparing the risk measures and the returns measures, which resulted in high rank correlations when I compared the performance measures*.

The second category comprises Treynor index, Appraisal ratio, and Generalized Sharpe ratio. Note that these indexes take into account the benchmark into the risk adjustment and are not considered in the work of Eling (2008). Correlations for this category are lower compared to the first one. Remember that risk adjustment in the first category is based on measures that capture systematic and non-systematic risk. In the second category we have measures that adjust returns in a different way. The Treynor index adjusts return by only systematic risk, for instance. So, the benchmark changes the way we look at risk and obviously we have another nature for risk adjustment.

The choice of one category of indexes instead of another one should be driven by the way investors look at risk. It means that if investors are concerned to performance in an absolute view, i.e. if investors neutral position is the risk-free rate, the first category of indexes should be more appropriate and the indexes in that category may result in similar conclusions regarding performance rankings. Nevertheless, if investors' neutral position is a benchmark portfolio, obviously we should choose performance measures of the second category for ranking purposes. See that the choice of the category may influence results.

The third category comprises the Upside Potential ratio and the MPPM. Note that the correlations are also lower than correlations in the first category. However, the explanation is different. See that numerator in the Upside Potential ratio is not the same as that of the first category. Furthermore, the risk adjustment intensity embedded in the risk aversion factor ρ in the MPPM has a nature different from the adjustment used in the indexes of the first category. So, the nature of the investors' utility function may influence ranking results.

Regarding the third category, if we are concerned to different reward functions, i.e. if we change the way we view performance and risk, or if we change the nature of the adjustment considering utility functions that change dramatically, we will change the relationship of risk and return.

The work of Eling (2008) does not consider in an appropriate manner the issues we have addressed here since he uses only performance measures of the first category. So, our results show that the way investors' perceive risk and return should affect the choice of the performance measure.

Therefore, the evaluation of mutual funds should not rely on a single performance measure. The use of different performance measures can bring robustness to results, and possibly avoid manipulation strategies with focus on specific measures. Another way to give robustness is to use data with different periodicity and time horizons. Also, qualitative evaluation of funds and managers can add value to the selection of mutual funds.

In a wider perspective, the use of a single or similar performance measures, like the ones from first category, is also detrimental for the funds industry dynamics and competition. The outperforming funds tend to increase their market share as clients usually have positive feedback preferences, acting as herds, and this induces managers to have very liquid portfolios. The use of a myriad of performance measures may reduce positive feedback flows since each investor would make his investment decisions considering a different basket of measures, and so reducing the herding effect.

DOI: 10.1002/ijfe

Moreover, under the perspective of the client, the use of various performance measures based on different assumptions would make clients more comfortable to adhere to an specific investment program which is able to match more risk and return dimensions relevant for him. As a consequence, it would also increase the attractiveness of the mutual fund's industry. Finally, the use of several performance measures, which lead to different rankings, tends to spread market shares over a greater number of funds and increase survivorship rates.

7. CONCLUSION

This paper evaluates empirically the ranking correlation of the Sharpe Ratio with 13 performance measures, for a sample of US mutual funds. The goal was to check whether the use of these measures alters significantly the ranking of mutual funds.

Our results indicate that when we use a set of similar performance measures, as the measures used by Eling (2008) and Eling and Schuhmacher (2007), the ranking correlations may be very high. But as we go further and expand the universe of performance measures, ranking correlations decrease.

Specifically, our results show that performance measures based on reward-risk ratios have similar rankings, when the numerator (mean excess return) is the same. This is in line with Eling (2008). However, when we move to other types of performances measures, results may be significantly different. This is the case of the MPPM, Upside Potential Ratio, and Appraisal Ratio. Results are especially different for the MPPM. Robustness checks show that some of performance measures are very sensitive to parameters' changes.

One weakness of this and other studies in this area is not to evaluate the statistical significance of the rankings. All funds are considered to have statically different performance measures. However, we cannot assure that a fund has a performance measure statistically higher than other fund without a more in-depth analysis of the statistical distribution of the measure. Thus, we would rank only with those funds that have in fact statistically different measures, and then compare rankings for different measures.

One robustness analysis that could be interesting to do is to change the time period of the study, and also include longer time periods. However, the survivorship issue present in our data set prevents us to do further analysis. Nevertheless, other studies used different time periods and performance measures, and did not find extremely high correlations as in Eling (2008).

Thus, our results for mutual fund data are in line with the work of Zakamouline (2009) for hedge fund data, i.e. the choice of performance measure is a relevant issue. Also, the paper of Cogneau and Hübner (2009b) found very low correlations for a small sample of 50 mutual funds. Therefore, we believe that the choice of the performance measure is actually important for mutual fund ranking and selection. The performance measure should depend on the investors' preferences and asset's behavior. We strongly recommend investors to make robustness analysis, through the use of different performance measures, time horizons, and time series periodicities.

We finally argue that the use of several performance measures has a positive impact on the mutual fund's industry, reducing the herding effect among investors, and so reducing the concentration of the industry.

ACKNOWLEDGEMENTS

The authors are grateful to Philippe Cogneau and an anonymous referee for their helpful comments.

NOTES

1. Goetzmann et al. (2007) argue that it should be selected to make holding the benchmark optimal for an uninformed manager. For the US market, this parameter would range from 2 to 4.
2. As pointed out by Noether (1981): 'The facts are that it is no easy matter to assign an operational interpretation to the Spearman coefficient. The Kendall coefficient, on the other hand, has an intuitively simple interpretation. What is more, its algebraic structure is much simpler than that of the Spearman coefficient. It can even be computed from the actual observations without first converting them to ranks.'

DOI: 10.1002/ijfe

REFERENCES

Agarwal V, Naik NY. 2004. Risk and portfolio decisions involving hedge funds. *Review of Financial Studies* **17**(1): 63–98.

Burke G. 1994. A sharper Sharpe ratio. *Futures* **23**(3): 56.

Cogneau P, Hübner G. 2009a. The (more than) 100 ways to measure portfolio performance. Part 1: standardized risk-adjusted measures. *Journal of Performance Measurement* **13**: 56–71.

Cogneau P, Hübner G. 2009b. The (more than) 100 ways to measure portfolio performance: Part 2: special measures and comparison. *Journal of Performance Measurement* **14**: 56–69.

Dowd K. 2000. Adjusting for risk: an improved Sharpe ratio. *International Review of Economics and Finance* **9**: 209–222.

Eling M. 2008. Does the measure matter in the mutual fund industry? *Financial Analysts Journal* **64**(3): 54–66.

Eling M, Schuhmacher F. 2007. Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking and Finance* **31**(9): 2632–2647.

Gehin W. 2004. A Survey of the Literature on Hedge Fund Performance. Available at SSRN: http://ssrn.com/abstract = 626441.

Goetzmann W, Ingersoll J, Spiegel M, Welch I. 2007. Portfolio performance manipulation and manipulation-proof performance measures. *The Review of Financial Studies* **20**(5): 1503–1546.

Gregoriou GN, Gueyie J-P. 2003. Risk-adjusted performance of funds of hedge funds using a modified Sharpe ratio. *Journal of Wealth Management* **6**: 77–83.

Henriksson RD, Merton RC. 1981. On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business* **54**: 513–533.

Hodges S. 1998. A generalization of the sharpe ratio and its applications to valuation bounds and risk measures, Working Paper, Financial Options Research Centre, University of Warwick.

Kaplan PD, Knowles A. 2004. A generalized downside risk-adjusted performance measure. Morningstar Associates and York Hedge Fund strategies (January).

Keating C, Shadwick W. 2002. A universal performance measure. *Journal of Performance Measurement* **6**: 59–84.

Kestner LN. 1996. Getting a handle on true performance. *Futures* **25**: 44–46.

Koekebakker S, Zakamouline V. 2008. Accounting for skewness preferences in investment decisions. FMA European Conference: Prague, Czech Republic.

Koekebakker S, Zakamouline V. 2009. Portfolio performance evaluation with generalized Sharpe ratios: beyond the mean and variance. *Journal of Banking and Finance* **33**: 1242–1254.

Malkiel B, Saha A. 2005. Hedge funds: risk and return. *Financial Analysts Journal* **61**(6): 80–88.

Morningstar. 2006. The Morningstar Rating^{TM} Methodology Morningstar Methodology Paper, Available at http://www.morningstar.com.

Noether GE. 1981 Why Kendall Tau? *Teaching Statistics* **3**(2): 41–43.

Shadwick WF, Keating C. 2002. A universal performance measure. *Journal of Performance Measurement* **6**(3): 59–84.

Sharpe WF. 1966. Mutual fund performance. *Journal of Business* **39**: 119–138.

Sharpe WF. 1994. The Sharpe ratio. *Journal of Portfolio Management* **21**: 49–58.

Sortino F, Van der Meer R. 1991. Downside risk. *Journal of Portfolio Management* **17**: 27–31.

Sortino F, Van der Meer R, Plantinga A. 1999. The Dutch triangle. *Journal of Portfolio Management* **26**(Fall): 50–58.

Stutzer M. 2000. A portfolio performance index. *Financial Analysts Journal* **56**(3): 52–61.

Treynor J. 1965. How to rate management of investment funds. *Harvard Business Review* **43**: 63–75.

Treynor JL, Black T. 1973. How to use security analysis to improve portfolio selection. *Journal of Business* 66–85.

Treynor J, Mazuy K. 1966. Can mutual funds outguess the market? *Harvard Business Review* **44**: 31–136.

Young TW. 1991. Calmar ratio: a smoother tool. *Futures* **20**(1): 40.

Zakamouline V. 2009. The choice of performance measure does influence the evaluation of hedge funds. Available at SSRN: http://ssrn.com/abstract = 1403246.

DOI: 10.1002/ijfe