Portfolio performance persistence: does the choice of performance measure matter?

January 14, 2019

Adcock, Christopher[*,†]; Areal, Nelson[‡,†]; Cortez, Maria Céu[‡,†]; Oliveira, Benilde[‡] and Silva, Florinda[§,‡,†]

Abstract

This paper investigates whether investment strategies using rankings based on different portfolio performance measures lead to different future abnormal returns. A set of 13 well known risk-adjusted performance measures is applied to a dataset of US equity mutual funds over the period July 1970 to March 2014. The results show some evidence of short-term performance persistence, suggesting that portfolios formed on different performance measures *ex ante* can generate abnormal returns *ex post*. A strategy of investing in the top performing funds and shorting the poor performing funds provides positive excess returns and five-factor alphas. However, when adjusting for the momentum factor, the abnormal performance mostly disappears. The results also show that overall there is little difference arising from the use of different performance measures, with one notable exception, the Rachev ratio.

Keywords: Downside risk measures; Portfolio performance measures; Performance rankings; *Ex post* performance; Portfolio performance persistence.

JEL classification: G11, G12.

Acknowledgements: This work was carried out within the funding with COMPETE reference nº POCI-01-0145-FEDER-006683, with the FCT/MEC's (Fundação para a Ciência e a Tecnologia, I.P.) financial support through national funding and by the ERDF through the Operational Programme on "Competitiveness and Internationalization – COMPETE 2020 under the PT2020 Partnership Agreement.

[*]School of Finance and Management, SOAS – University of London, UK

[†]NIPE research center - School of Economics and Management, University of Minho, Braga, Portugal

[‡]School of Economics and Management, University of Minho, Braga, Portugal

[§]Corresponding author; email: fsilva@eeg.uminho.pt; address: Universidade do Minho - Escola de Economia e Gestão, Gualtar 4710-057 Braga-Portugal

1 Introduction

Portfolio performance evaluation is an important and highly debated subject in the finance literature. Since the advent of modern portfolio theory and the Capital Asset Pricing Model (CAPM) in the 1960s, many portfolio performance measures have been proposed in the literature. Yet, there is no consensus on what risk measure is most appropriate to be incorporated in performance measures and a discussion on whether the choice of such measures matters or not is ongoing.

Traditionally, it is considered that use of the ubiquitous Sharpe ratio (Sharpe, 1966) as the performance measure of choice requires that returns have a normal distribution, investors have a quadratic utility function or otherwise accept volatility as the measure of risk. More recent research, however, indicates that the Sharpe ratio is the appropriate performance measure under much more general conditions, namely that returns follow an elliptically symmetric distribution. Nonetheless, traditional performance measures do not explicitly take into account the risk premia due to all higher moments or co-moments. Research that has analyzed this issue may be divided into two groups. First, several studies have extended standard portfolio theory to incorporate the effect of skewness (Arditti, 1967; Arditti and Levy, 1975; Kraus and Litzenberger, 1976). Harvey and Siddique (2000) develop an asset pricing model with conditional coskewness where risk averse investors prefer positively skewed assets to negatively skewed assets. Assets that decrease a portfolio's skewness are less desired and thus should have higher expected returns[5]. Additionally, Dittmar (2002) finds a preference for lower kurtosis. Assets that increase a portfolio's kurtosis are less desirable and should also have higher expected returns. A second group of papers considers asymmetry in returns and focuses on risk measures derived from the tails of the distribution. These measures have come to be known as downside risk measures. Downside risk-adjusted performance measures are considered theoretically more robust: they do not assume normality of returns and do not rely on volatility as a measure of risk. They are widely used in portfolio performance evaluation.

Recent studies have questioned whether the choice of performance measure (traditional or downside risk-adjusted) matters in practice. The empirical evidence is controversial. Some studies show that the use of different risk-adjusted performance measures does not seem to matter for investment fund decisions based on performance ranked portfolios (Auer, 2015; Eling, 2008; Eling and Faust, 2010; Eling and Schuhmacher, 2007). This is because the empirical correlation between performance measures and the Sharpe ratio is equal to or close to unity. Thus, there is no difference in fund rankings resulting from the use of different performance measures. Others studies document that it does matter at least for some of the performance measures under consideration (Adcock et al., 2017; Carles et al., 2018; Ornelas et al., 2012; Zakamouline, 2011). In particular,

[5] Ding and Shawky (2007), Kostakis (2009) and Moreno and Rodríguez (2009) are examples of studies that consider the Harvey and Siddique (2000) model to evaluate fund performance.

both Zakamouline (2011) and Ornelas et al. (2012) point out that fund rankings can vary substantially depending on the performance measure used. In addition, as first pointed out by Zakamouline (2011), many performance measures are monotonic functions of the Sharpe ratio when computed using formulae based on the normal distribution. In such cases, the correlation between a performance measure and the Sharpe ratio will be unity or close to it (see for example Adcock et al., 2012). Furthermore, Schuhmacher and Eling (2012), Adcock et al. (2012, 2017) and Guo and Xiao (2016) show that for a broad class of probability distributions of portfolio returns, many performance measures that have been reported in the literature are also monotonic functions of the Sharpe ratio. Distributions that exhibit asymmetry as well as symmetric forms of non-normality are included in this class. However, as reported in Zakamouline (2011) and Adcock et al. (2017), monotonicity with respect to the Sharpe ratio can fail if returns follow different distributions. From a practical perspective, this suggests that the usefulness of a given performance measure for a set of funds depends on the homogeneity of the distributions of returns.

The implications of the debate summarized in the previous paragraph are of significant practical value to investors, who typically pay great attention to fund performance rankings in their investment decisions (Capon et al., 1996), implicitly assuming performance persistence. Nevertheless, if rankings of funds change according to the performance measure used, portfolios formed on the basis of these rankings may lead to different future performance. This is an important issue that requires further research. The purpose of this paper is therefore to investigate whether there is any performance measure that is better able to detect performance persistence. Additionally, we investigate whether the different *ex ante* performance measures generate different future portfolio performance. Although there is extensive empirical evidence addressing the correlations between different performance measures, these analyses are restricted to a static approach, thereby ignoring the possibility of time-varying correlations. If the correlations between the different performance measures do vary over time, the use of alternative measures may impact *ex post* performance results. By simulating investment strategies over time, our analysis captures the dynamic nature of the correlations. Whether this impact is relevant or not is mainly an empirical issue that this paper addresses.

The dataset consists of portfolios of US domestic equity funds over the period July 1970 to March 2014. Each month portfolios of funds are created based on rankings of funds that result from applying a number of different risk-adjusted performance measures. The *ex post* performance of these portfolios of funds is measured using excess returns, Carhart (1997) alphas and Fama and French (2015) alphas. In addition to the Sharpe ratio, 12 other well known risk-adjusted performance measures are analyzed. Most of the previous studies that investigate whether past performance is related to future performance use a single or a very limited set of performance measures to rank funds. For instance, Hendricks et al. (1993), Brown and Goetzmann (1995), Elton et al. (1996, 2012), Carhart (1997) and Bollen and Busse (2005) document evidence of performance persistence when ranking funds on the basis of raw returns and abnormal returns (alphas). However, evidence of persistence is equivocal. Some authors find that performance persistence is short-lived (see for example Hendricks et al., 1993 and Bollen and Busse, 2005) while others observe performance persistence over periods longer than one year (Elton et al., 1996, 2012). Furthermore, Carhart (1997) attributes the persistence phenomenon to the momentum effect and observes that persistence is concentrated in the poor performing funds.

We are not aware of studies that investigate performance persistence using downside-risk based measures to rank fund performance. To the best of our knowledge this is also the first study exploring the persistence of mutual fund performance in the context of a range of performance measures, including downside risk-adjusted measures and also a set of measures commonly used by practitioners. In addition to the detailed results based on 13 risk-adjusted performance measures reported in this paper, a set of results covering 82 performance measures is available on request.

The paper is structured as follows. In section two the risk-adjusted performance measures are listed and the methodology used to assess *ex post* fund performance is described. Section three presents the data. Section four reports and discusses the empirical results. Section five concludes. Additional detailed supporting results are available from the corresponding author.

2 Methods

2.1 Risk-adjusted performance measures used in ranking funds

We apply a set of risk-adjusted performance measures to analyze if investment strategies using rankings based on these measures lead to different *ex post* performance results. Each period fund performance is assessed using different risk-adjusted measures. The use of some of these performance measures is strongly motivated by the finance literature and others are commonly used in practice by fund managers. The most known risk-adjusted performance measure and probably the most used in practice is the traditional Sharpe ratio (Sharpe, 1966). In the context of modern portfolio theory, there are also strong theoretical arguments for the use of the Appraisal ratio as a measure of performance (Treynor and Black, 1973). More recently, Goetzmann et al. (2007) develop an alternative to the traditional Sharpe ratio that they denominate as the Manipulation Proof Performance Measure. As this measure captures the whole profile of the returns distribution, Goetzmann et al. (2007) claim that it is immune to fund manager manipulation that may occur when performance measures based only on the first two moments are used. General difficulties associated with the use of the traditional performance measures in the context of non-elliptically symmetric distributed returns also motivated the development of an alternative downside risk measurement framework. In addition, the early 1990's witnessed an intensification of the ongoing debate about good risk management practices. As a consequence, several alternative downside risk-adjusted performance ratios were proposed. These ratios are mainly based on the Lower Partial Moments (like the Modified Sortino and Farinelli-Tibiletti ratios) and on the Value at Risk (Excess Return on Value at Risk, Excess Return on Conditional Value at Risk and Rachev ratio). Over the years a set of performance measures based on drawdown (Calmar ratio, Sterling ratio and Burke ratio) became very popular among practitioners and therefore they are also often included in studies that aim the comparison of different measures of performance. A brief description of each of these portfolio performance measures is presented below. The list and accompanying descriptions use the nomenclature in Zakamouline (2011), which is also reported in Adcock et al. (2012).

Sharpe ratio (Sharpe, 1966)

$$\text{Sharpe Ratio} = \frac{(\mu - r_f)}{\sigma}, \tag{1}$$

where μ represents the portfolio average returns, r_f is the risk free rate and σ represents the standard deviation of the returns.

Appraisal ratio (Treynor and Black, 1973)

$$AR = \frac{\alpha}{\sigma_\epsilon}, \tag{2}$$

where α measures the abnormal return using the CAPM and σ_ϵ measures the unsystematic risk proxied by the standard deviation of the model residuals.

Manipulation Proof Performance Measure (Goetzmann et al., 2007)

$$MPPM = \{\Delta t(1-\omega)\}^{-1} log\{T^{-1} \sum_{t=1}^{T} e^{(1-\omega)R_t}\}, \tag{3}$$

where $\{R_t\}$ represents a time series of fund returns, the scalar ω denotes risk aversion and Δt is the length of the time interval. Following both Goetzmann et al. (2007) and Brown et al. (2010), a risk aversion coefficient of 3 is used as it is considered to be representative of institutional investors.

The next two measures are based on Value at Risk (VaR) and Conditional Value at Risk (CVaR). VaR is generally defined as the maximum expected loss over a given horizon period at a given probability of $1 - \alpha$. CVaR is the expected loss under the condition that VaR is exceeded.

Expected excess return to Value at Risk (Dowd, 2000)

$$\text{ERVaR} = \frac{(\mu - r_f)}{VaR_\alpha}. \tag{4}$$

Expected excess return to Conditional Value at Risk (Martin et al., 2003)

$$\text{ERCVaR} = \frac{(\mu - r_f)}{CVaR_\alpha}. \tag{5}$$

In this paper VaR and CVaR are computed parametrically assuming a Normal-GARCH model and also non-parametrically using historical returns and numerical integration. In practice, VaR and CVaR estimates are computed from a 10% to 0.1% loss probability (α), but the most commonly used probabilities are 5% and 1%. For monthly data we consider that a probability of 5% is generally adequate.

The Rachev ratio is the conditional expected excess return in the right hand tail of the distribution, divided by the conditional expected excess return in the left hand tail of the distribution.

Rachev ratio (Rachev et al., 2007)

$$RR(\alpha, \beta) = \frac{\text{CVaR}_\beta(r_f - x)}{\text{CVaR}_\alpha(x - r_f)}. \tag{6}$$

In this paper the RR is computed non-parametrically using the empirical distribution of the returns. Following Zakamouline (2011), a combination of $\alpha = 0.05$ and $\beta = 0.05$ is used to compute the ratio.

Next, several performance ratios based on upper and/or lower partial moments are presented. Both lower and upper partial moments are defined with respect to a reference point τ which may take any real value

$$L_\omega(\tau) \quad = \int_{-\infty}^{\tau} (\tau - x)^\omega f(x) dx \tag{7}$$

$$H_\omega(\tau) \quad = \int_{\tau}^{\infty} (x - \tau)^\omega f(x) dx, \tag{8}$$

where $f(.)$ denotes the probability density function of portfolio returns. It is usually assumed that $\omega \geq 1$. In practice usually $\omega = k$ is an integer, typically equal to 1 or 2.

Modified Sortino ratio (Pedersen and Satchell, 2002)

$$SP = \frac{(\mu - r_f)}{\sqrt{L_2(r_f)}}, \tag{9}$$

where L_2 is the lower partial moment of order 2.

Farinelli-Tibiletti ratio (Farinelli and Tibiletti, 2008)

$$FT(\alpha,\beta)=\frac{\{H_\alpha(r_f)\}^{\frac{1}{\alpha}}}{\{L_\beta(r_f)\}^{\frac{1}{\beta}}};\ \alpha,\beta\geq 1, \tag{10}$$

where H_α and L_β are, respectively, the upper and lower partial moments of any order α and β. There are other measures that are special cases of the FT ratio. The Omega measure (Keating and Shadwick, 2002) and Kappa measure (Kaplan and Knowles, 2004) of order one plus one are equal to $FT(1,1)$. The Upside potential ratio (Sortino et al., 1999) is equivalent to $FT(1,2)$. By considering $FT(1,1)$ and $FT(1,2)$, we are also able to make inferences for those measures.

Finally, three measures most commonly used in practice and that are based on drawdown, are presented.

Calmar ratio (Young, 1991)

$$\text{Calmar}=\frac{\mu-r_f}{-\text{MD}_1}, \tag{11}$$

where MD_1 represents the lowest return in a given time series of returns.

Sterling ratio (Kestner, 1996)

$$\text{Sterling}=\frac{\mu-r_f}{\frac{1}{N}\sum_{i=1}^{N}-\text{MD}_i}, \tag{12}$$

where N represents some predefined set of the largest drawdowns.

Burke ratio (Burke, 1994)

$$\text{Burke}=\frac{\mu-r_f}{\sqrt{\sum_{i=1}^{N}\text{MD}_i^2}}. \tag{13}$$

Similarly to Eling and Schuhmacher (2007), N is set equal to 5 for the computation of both the Sterling and the Burke ratios.

Overall, 13 performance measures are computed and analyzed in detail. These measures are part of a larger set of 82 performance measures that were computed, and whose results are available upon request[6].

[6] The additional measures consider probabilities of 0.1%, 0.5%, 1% and 2.5%, all combinations of orders 1 to 5 for the FT ratios, three additional AR, based on the three-factor model of Fama and French (1993), the four-factor model of Carhart (1997) and the five factor model of Fama and French (2015) and the MPPM with risk aversion coefficients of 2 and 4.

2.2 Constructing and evaluating portfolios

We analyze equally weighted and value weighted portfolios of mutual funds ranked according to different risk-adjusted performance measures, as described in the previous section. Every month we create portfolios based on deciles of funds ranked according to their past performance – the rank portfolios. An estimation window of 60 months (5 years) is used to compute estimates of performance. In each period the oldest observation is dropped and the most recent observation is added. For each rank portfolio we end up with a time series of monthly portfolio returns (over the period July 1970 to March 2014).

The *ex post* performance of these rank portfolios is evaluated by computing excess returns, the four-factor alpha (Carhart, 1997) and the five-factor alpha (Fama and French, 2015). The motivation for using different performance measures in the ranking procedure and in the *ex post* performance assessment follows from Carhart's (1997) argument that using different measures to sort funds and estimate *ex post* performance avoids possible model biases. In addition, by controlling for additional sources of systematic risk, the four- and five-factor model alphas capture solely the returns associated with fund managers' skills. It is worth noticing that the reward-to-risk ratios used ex-ante do not control for multiple sources of risk and can seemingly generate the appearance of abnormal performance that is simply due the returns associated to the omitted factor loadings. Considering the Carhart (1997) four factor model, the performance measure (alpha) is obtained by the following equation

$$r_{p,t} - r_{f,t} = \alpha_p + \beta_p(r_{m,t} - r_{f,t}) + s_p SMB_t + h_p HML_t + p_p MOM_t + \epsilon_{p,t}, \qquad (14)$$

where $r_{p,t}$ represents fund returns, $r_{f,t}$ is the risk free rate, α_p represents the fund performance measure, $r_{m,t}$ represents market returns, SMB_t, HML_t and MOM_t represent the size, value and momentum factors respectively. SMB_t is the return on a diversified portfolio of small stocks minus the return on a portfolio of big stocks, HML_t is the return difference between portfolios of value stocks and growth stocks, and MOM_t is the return difference between high prior return portfolios and low prior return portfolios.

The performance measure based on the Fama and French (2015) five factor model results from

$$r_{p,t} - r_{f,t} = \alpha_p + \beta_p(r_{m,t} - r_{f,t}) + s_p SMB_t + h_p HML_t + r_p RMW_t + c_p CMA_t + \epsilon_{p,t}, \qquad (15)$$

where RMW is the profitability factor, measured as the difference between the returns on diversified portfolios of stocks with robust and weak profitability and CMA is the investment factor, measured as the difference between the returns on diversified portfolios of stocks of low and high investment firms. The other variables are as defined above.

We also compute two types of differences portfolios. The first one corresponds to the difference in returns between the highest performance decile and the lowest performance decile considering the same risk-adjusted performance measure. This differences portfolio simulates the performance that results from a strategy of buying the top decile and shorting the bottom decile portfolio (long-short strategy), thereby assessing the profitability of exploring a performance persistence investment strategy. The second one corresponds to the difference in returns between the different performance measures within the same decile portfolio. The aim is to assess if the choice of a specific performance measure, particularly at each of the extreme deciles (deciles 1 and 10), leads to different *ex post* performance.

3 Data

The dataset consists of US domestic equity funds of different categories obtained from the CRSP database. Returns with a monthly frequency are available from December 1961 onwards. Only funds with an objective code of EQC (Equity, Domestic and Cap-based) and EQY (Equity, Domestic and Style) are considered. All funds classified as ETFs, ETNs and Index funds are excluded. Funds with less than 60 months of observations are also eliminated. In the case of funds with multiple share classes, only one is considered: the one with the longest historic record and, if necessary, the one with the largest Total Net Assets (TNA). To avoid the incubation bias (Evans, 2010)[7] and the omission bias (Elton et al., 2001)[8] the first three years of data and funds with less than $15 million in TNA, respectively, are removed. This leads to a final dataset of 2238 funds. Figure 1 reports the number of funds with 60 months of history over the period December 1969 to March 2014. As can be observed, there is a considerable increase in the number of funds throughout this period.

[Insert figure 1 here]

The first decile portfolios are formed in June 1970. Funds that disappear during a particular month are included in the portfolios until they disappear and then the portfolios weights are readjusted accordingly. The *ex post* performance of these decile portfolios is then assessed considering the period July 1970 to March 2014 (a total of 525 monthly return observations). Fund returns are net of management expenses and

[7] Evans (2010) shows that there is an incubation bias in the CRSP mutual fund database. When funds are included in the database for the first time, they bring all their past history (even if they were being privately traded). Since only the most successful incubator funds will be publicly traded, this creates a bias for the incubation period. Evans (2010) shows that an age filter effectively removes this bias.

[8] The omission bias arises due to the different frequencies of returns available on the CRSP database (e.g. monthly or annual returns). To avoid this problem Elton et al. (2001) restrict the sample of funds to contain only those funds that have over $15 million in TNA at the beginning of any observation period as CRSP reports monthly data for most funds with over $15 million in TNA.

security-level transaction costs, but do not include load fees. The monthly risk free rate as well as the monthly returns of the risk factors (market, size, value, momentum, profitability and investment factors) over the same period were downloaded from the Professor Kenneth French webpage.

4 Empirical results and discussion

4.1 Performance persistence with monthly portfolio rebalancing

In this section we analyze whether there is any performance measure that is better able to detect performance persistence. The results consider a total of 13 different risk-adjusted performance measures, as described in section 2.1. As previously mentioned, the ERVaR and the ERCVaR are computed parametrically assuming a Normal-GARCH model[9] and non-parametrically by using historical simulation; that is, using the empirical distribution of past returns.

We start by analyzing equally weighted decile portfolios formed on the basis of the different risk-adjusted performance measures. In each month, funds are ranked according to their performance over the previous 60 months and decile portfolios of funds are formed. These portfolios are then rebalanced monthly. Decile 1 corresponds to the portfolio of the bottom performing funds and decile 10 corresponds to the portfolio of the top performing funds. Value weighted decile portfolios are also analyzed. *Ex post* portfolio performance is assessed using excess returns over the risk-free rate and the benchmark, as well as alphas based on the Carhart (1997) and Fama and French (2015) models.

Table 1 reports the mean excess returns over the market benchmark[10] for the equally weighted decile portfolios over the period from July 1970 to March 2014[11]. The results show that the excess returns are all positive and statistically significant (at the 1% significance level), ranging from 0.610% (decile 1) and 0.932% (decile 10) monthly excess returns (or 7.3% and 11.2% annualized excess returns). For most risk-adjusted performance measures there is a tendency for the highest performance deciles to present the highest excess returns. The exception is observed mainly in relation to the RR, for which there is no clear tendency of higher or lower excess returns across different deciles. In

[9]When VAR and CVaR are cumputed parametrically, assuming a Normal-GARCH model, ERVaR and ERCVaR produce exactly the same rankings and therefore only the results for ERVaR (ERVaR-NG) are reported.

[10]As the market benchmark we considered the US market return available on the Professor Kenneth French webpage.

[11]We also analyzed other descriptive statistics for the decile portfolios excess return series. All the decile portfolios exhibit negative skewness and positive excess kurtosis. In all cases we reject the hypothesis of a normal distribution according to the Jarque-Bera test.

each decile excess returns are similar for the majority of the performance measures (11 out of the 13). The two exceptions are the RR and to a lesser extent the MPPM. Looking at the unreported statistics of the excess returns over the risk free rate, the results are alike.

[Insert Table 1 here]

Tables 2 and 3 present the *ex post* performance of the equally weighted decile portfolios estimated by the Carhart (1997) four factor and the Fama and French (2015) five factor models, respectively. The results on the four factor model (Table 2) show that decile 1 (the lowest performance portfolio) alphas are mostly negative and statistically significant (at the 5% level), ranging from -0.172% and -0.082%. We also observe a tendency for alphas to increase from the bottom performing decile to the top performing decile, with the exception of the RR, for which there is no clear tendency and even a slightly more negative alpha for the top performing decile (-0.084%) than for the bottom performing decile (-0.082%) is observed. The fact that the decile portfolios' performance increase almost monotonically is consistent with the existence of persistence of alphas. All alphas of decile 10 are neutral, and excluding the one obtained using the RR, they vary between -0.028% and 0.047%. With respect to the five-factor alphas (Table 3) the results are similar. Decile 1 alphas are all negative and statistically significant, ranging from -0.240% and -0.107%, while decile 10 alphas are all positive but not statistically significant, ranging from 0.026% and 0.096%. It is worth pointing out that the evidence of underperformance in decile 1 is stronger than that observed with the four factor model as most of the negative alphas are statistically significant at the 1% level. Interestingly, the monotonicity of the alphas with respect to the deciles is stronger numerically for the four factor model than the five factor model.

[Insert Table 2 here]

[Insert Table 3 here]

Unreported results on the *ex post* performance of value weighted portfolios are similar in terms of excess returns, either over the market benchmark and the risk free rate. In relation to *ex post* alphas, those of the bottom portfolios are clearly less negative, suggesting that the negative performance of the bottom decile portfolios may be driven by smaller size funds[12].

[12] To further investigate the relationship between portfolio performance and the ranking of the funds, we run regressions of *ex post* portfolio performance on the different deciles as follows $P_{p,d} = a + \beta d + \epsilon_{p,d}$, where $P_{p,d}$ is the *ex post* performance from *ex ante* performance measure p in decile d. We have also allowed the intercept of the regression (a) to vary according to *ex ante* performance measures. The results show that there is a positive and statistically significant relation between *ex post* performance, measured either by excess returns or four-factor and five-factor alphas, and the deciles. As expected, when we exclude the portfolios based on the RR from this analysis both the slopes of the regressions and the adjusted R-squareds increase.

Table 4 presents excess returns, four-factor alphas, and five-factor alphas of a portfolio that corresponds to the differences between the top (10) and bottom (1) equally weighted and value weighted deciles. This portfolio shows the returns that result from a strategy of buying the top decile funds and shorting the bottom decile funds. As can be observed, we find positive and statistically significant (at the 5% level) excess returns and alphas in most of the differences portfolios formed on the basis of different measures of performance. In the case of the differences portfolios formed on the RR and MPPM there is far less evidence of statistically significant alphas. These results seem to suggest that the RR and the MPPM do the worst job in predicting future performance. Furthermore, the evidence in favor of an outperformance of an investment strategy exploiting performance persistence is stronger when we use returns and the five-factor alphas to assess *ex post* performance, with most values exhibiting statistical significance at the 1% level. However, the evidence of abnormal returns is scarce when the four factor model is used either for equally or value weighted portfolios. For instance, in the case of value weighted portfolios we find evidence of performance persistence (at the 5% level) only for two of the measures (ERVaR-NG and AR-CAPM). The weaker evidence of a profitable strategy exploiting performance persistence observed in the context of the Carhart (1997) four factor model suggests that abnormal returns are mainly driven by short-term momentum and raises the issue of whether the five factor model of Fama and French (2015) does a good job in evaluating performance[13]. Analysing in more detail the regressions estimates we observe that the explanatory power of both models, measured by the R-squareds, is similar and always above 90%. In the case of the four factor model, the size and momentum are relevant risk factors. The size factor is statistically significant across the different deciles and the different performance measures. The momentum factor is positive and statistically significant for the top performing deciles, whereas it is negative in some of the bottom performing deciles. In relation to the five factor model the most relevant factors are size and profitability.

[Insert Table 4 here]

4.2 Implications of using different performance measures

Besides analyzing the *ex post* performance of an investment strategy that consists in selecting funds according to a specific performance measure, it is also relevant to investigate whether using different measures of performance to rank funds matters. In this section, we focus on the latter issue by assessing whether there are differences in the performance of decile portfolios formed on the basis of rankings of different risk-adjusted performance measures. For each pair of performance measures, a portfolio that corresponds to the return differences is formed. Then, the *ex post* performance of the differences portfolio

[13]See for example Barillas and Shanken (2018) for a comparison on alternative asset pricing models. They show that the Fama and French (2015) five factor model is dominated by models that include a momentum factor.

for each pair of measures is computed using, as before, excess returns, four-factor alphas and five-factor alphas. Tables 5 to 8 report estimates of performance differences of equally weighted portfolios (performance differences between portfolios formed using measure in column minus measure in line). In the sake of brevity, we only report the results for the extreme decile portfolios (decile 1 represents the bottom performing funds and decile 10 represents the top performing funds) and based on excess returns and four-factor alphas[14]. The entries in these tables are derived from the corresponding cells in tables 1 and 2.

Tables 5 and 6 report the results based on excess returns for deciles 1 and 10, respectively. For the majority of the pairs of performance measures we do not observe statistically significant differences (at the 5% level) in *ex post* excess returns. Only the use of the RR seems to lead to significant return differences. For decile 1, the RR tends to yield higher returns while for decile 10 it seems to generate lower returns. This seems to indicate that this measure is not able to correctly identify the best and the worst performing funds.

[Insert Table 5 here]

[Insert Table 6 here]

Tables 7 and 8 show the *ex post* performance differences (measured by four-factor alphas) for each pair of performance measures for deciles 1 and 10, respectively. In the case of decile 1, with one exception (the pair AR-CAPM, RR), all the reported alphas are insignificant (at the 5% level). For decile 10 some pairs of performance measures show statistically significant alphas. In general, the systematic use of the RR, in comparison with the other performance measures, seems to lead to worse *ex post* performance.

[Insert Table 7 here]

[Insert Table 8 here]

Although not reported, we also analyze the *ex post* performance differences for each pair of performance measures for value weighted portfolios. In general, the results are similar in what concerns the fact that most of the alphas for the differences portfolios are not different from zero. The majority of the performance measures under analysis seem to lead to similar future performance results. When statistically significant differences do occur, in general they are related with the use of the RR, the MPPM, the Calmar and the Burke ratios. However, the impact of using these performance measures is not clear, as the reported differences vary between positive and negative.

[14] The results across the other deciles are similar.

4.3 Portfolio turnover and the impact of transaction costs

An important issue when analyzing the long-short investment strategy described above (buying top performers and shorting bottom performers) is the frequency of trading and the impact of transaction costs. In this section we investigate whether the evidence of short term persistence can be exploited in practice by investors. For this purpose, portfolio turnover is computed and the impact of transaction costs on portfolio performance is analyzed. As our analysis focuses on portfolios of mutual funds we need to take into account possible load fees they might charge. Load fees differ across funds. Some funds have front end fees and/or back end fees. Usually, front end fees are waived and back end fees vary according to the investment horizon. Considering the typical load fees structure of US equity funds, we analyze the impact of transaction costs of 0.25% and 2%. For each portfolio, transaction costs (c) at time t are equal to

$$c_t = c \sum_{t=1}^{N} |w_{i,t} - w_{i,t^-}|, \tag{16}$$

where $\sum_{t=1}^{N} |w_{i,t} - w_{i,t^-}|$ is the portfolio turnover[15].

Table 9 reports the turnover of the equally weighted portfolios. Monthly turnover ranges from around 8% to 15%. In general, the levels of portfolio turnover are lower in the extreme deciles (with decile 10 exhibiting the lowest figures) and higher for the middle ones. The same pattern is observed for value weighted porfolios, although in this case the levels of turnover are higher, ranging between 12% and 30%. Anyhow, in any decile, the turnover does not vary much as a result of performance measure choice. Even the RR leads to similar values of this indicator. We also observe that some decile portfolios formed on the basis of ERVaR-NG have a slightly higher turnover.

[Insert Table 9 here]

The net returns of the long-short strategy are computed as the difference in the returns of the best (decile 10) and worst (decile 1) performing portfolios minus the sum of the transaction costs associated with both portfolios. Table 10 shows the differences in performance for equally weighted portfolios that arise from this long-short strategy considering transaction costs of 0.25% and 2%. As expected, *ex post* performance decreases as transactions costs increase. For low values of transactions costs of 0.25%, the results show that investors can profit from exploiting persistence strategies in terms of

[15]Portfolio turnover is computed as follows. On month $t-1$ the *ith* fund has a given weight in the portfolio denoted by $w_{i,t-1}$, with $i = 1, ..., N$. The portfolio return is $r_{p,t} = \sum_{t=1}^{N} w_{i,t-1} r_{i,t}$, where $r_{i,t}$ is the return on the *ith* fund on month t. At moment t^-, just prior to rebalancing, the actual weight of the *ith* fund in the portfolio is $w_{i,t^-} = w_{i,t-1} \frac{1+r_{i,t}}{1+r_{p,t}}$. The required rebalancing at time t is equal to $w_{i,t} - w_{i,t^-}$, where $w_{i,t}$ is the new weight for fund i using time t information.

excess returns and five-factor alphas. However, the abnormal returns disappear when the momentum factor is accounted for. With transactions costs of 2%, long-short strategies based on any of the *ex ante* performance measures lead to unprofitable results whatever performance measure is considered. Similar conclusions hold for value weighted portfolios.

[Insert Table 10 here]

4.4 Alternative portfolio cut-offs and rebalancing periods

For robustness purposes we also form portfolios with different cut-offs: considering the top 30%, the middle 40% and the bottom 30% performing funds. The results in Table 11 show that the differences between top performing and bottom performing funds are lower compared with those obtained with the decile portfolios, as expected. Compared with the results of Table 4, there is also slight less evidence of performance persistence for equally weighted portfolios and almost no evidence of persistence in the case of value weighted portfolios when using alphas from the four factor model.

[Insert Table 11 here]

So far, the results considered monthly rebalancing of the portfolios. We also analyze the results obtained with an annual rebalancing strategy. Each year (end of June), funds are ranked according to their performance over the previous 60 months and decile portfolios of funds are formed. The composition of the portfolios is then maintained until June of the following year. The *ex post* performance results (excess returns, four-factor and five-factor alphas) of the differences portfolios between the top (decile 10) and bottom (decile 1) performing funds, considering annual rebalancing, are reported in Table 12.

[Insert Table 12 here]

Comparing the results in Table 12 with those reported in Table 4, we observe lower excess returns and alphas and clearly much less evidence of abnormal returns from investment strategies exploiting performance persistence. In fact, abnormal returns (at the 5% level) hold mainly for the equally weighted portfolios and when performance is measured with the four-factor alpha. For value weighted portfolios only ERVaR-NG seems to exhibit predictive ability of future performance. These findings suggest that performance persistence is more concentrated in smaller size funds. The fact that the five-factor alphas are no longer statistically significant confirms that the evidence in favor of performance persistence for monthly rebalancing with this model is mostly due to the short-term momentum effect.

With regards to the question of whether using different risk-adjusted performance measures to rank funds is relevant, the unreported results for the extreme portfolios (deciles 1 and 10) are similar to those obtained with monthly rebalancing, as in most cases there are no statistically significant differences in performance for any pairs of performance measures. The few cases in which such differences exist are associated with the use of RR, AR-CAPM, MPPM and FT(1,1).

To further explore these results we also analyze the *ex ante* correlations over time between the Sharpe ratio and the other performance measures. For each year (in June), we test the null hypothesis that the correlation between the Sharpe ratio and other performance measures is equal to one (Adcock et al., 2012) and also analyze the correlation between the rankings produced by these performance measures. Unreported results suggest that some performance measures are more stable than others as far as the correlation with the Sharpe ratio is concerned. The null hypothesis of unit correlation is rejected in all the years for the RR. With regards to the other measures, the rejection of the null occurs occasionally. For the ranking correlations, the unreported results show changes in rank ordering, which is consistent with the findings reported by Zakamouline (2011) and Ornelas et al. (2012). It is worth mentioning that the RR also presents the lowest rank correlation with the Sharpe ratio[16]. The variability in the *ex ante* correlations implies that fund selection will vary depending on the performance measure used. Overall, however, the *ex post* performance is not that different for most of the performance measures.

5 Conclusions

Several papers on the performance of investment portfolios report mixed results on whether the choice of performance measure matters. In this paper, we investigate the extent to which the *ex ante* use of a set of 13 well-known performance measures generates abnormal *ex post* performance and whether the performance measure used makes any difference. The study is based on a dataset of monthly returns of US equity mutual funds from 1970 to 2014.

For each risk-adjusted performance measure, we analyze the *ex post* performance of a strategy of holding decile portfolios formed using fund rankings on that performance measure. We also investigate the results of a long-short strategy of buying top decile and selling the bottom decile funds, thus simulating an investment strategy exploiting persistence in fund performance. The estimates of the performance measures and hence the constituents of the decile portfolios are recomputed each month. The results show that decile portfolio performance increases almost monotonically from the worst decile

[16]The low correlation between these measures has been documented in some previous studies (Adcock et al., 2017; Eling et al., 2011), although the reasons pointed out by the authors for that do not fully explain our results.

to the best. This indicates that selecting the best funds according to the *ex ante* values of a performance measure leads to higher performance *ex post*. The results of the long-short investment strategy show that investors exploiting a persistence strategy can obtain abnormal returns. However, when adjusting for the momentum risk factor, this outperformance disappears. These results suggest that short term performance persistence is driven by the momentum effect. It is also worth mentioning that the measure based on normal GARCH risk estimates seems to consistently provide more evidence of persistence. When considering trading costs, the abnormal excess returns and five-factor alphas are still observed, but only for the lowest level of transaction costs (0.25% round trip).

The paper also presents a detailed analysis of the effect of different performance measures on *ex post* performance. In general, the results show that the use of alternative performance measures leads to very similar *ex post* performance. The main exception is the Rachev ratio, which is unable to discriminate between the performance of funds in the dataset used in this study. Portfolios formed using the bottom and top 30% rule and portfolios formed with annual rebalancing show less evidence of performance persistence, in particular for value weighted portfolios.

In sum, the results of this study suggest that *ex post* there is no reason to use any performance measure other than the Sharpe ratio, even though *ex ante* there are often considerable differences in the rankings. If anything, the results of the ERVaR-NG may support using conditional volatility in a alternative form of the SR. This finding thus complements those studies that report high correlation, often values of unity or close to it, of performance measures with the Sharpe ratio. As this is an empirical study, it is appropriate to note that different results could arise for different datasets of asset or fund returns. It is well known, for example, that high or even unit correlations with the Sharpe ratio arise as a result of theoretical properties of the return distributions. These properties may be satisfied by US mutual funds collectively, but would not necessarily be met by other datasets.

References

Adcock, C., N. Areal, M. Armada, M. C. Cortez, B. Oliveira, and F. Silva (2012). "Tests of the correlation between portfolio performance measures". *Journal of Financial Transformation* 35, pp. 123–132.

— (2017). "Portfolio performance measurement: monotonicity with respect to the Sharpe ratio and multivariate tests of correlation". URL: https://ssrn.com/abstract=2707523.

Arditti, F. D. (1967). "Risk and the required return on equity". *The Journal of Finance* 22 (1), pp. 19–36.

Arditti, F. D. and H. Levy (1975). “Portfolio efficiency analysis in three Moments: the multiperiod case”. *The Journal of Finance* 30 (3), pp. 797–809.
Auer, B. R. (2015). “Does the choice of performance measure influence the evaluation of commodity investments?” *International Review of Financial Analysis* 38, pp. 142–150.
Barillas, F. and J. Shanken (2018). “Comparing Asset Pricing Models”. *The Journal of Finance* 73 (2), pp. 715–754.
Bollen, N. P. and J. A. Busse (2005). “Short-term persistence in mutual fund performance”. *The Review of Financial Studies* 18 (2), pp. 569–597.
Brown, S. J. and W. N. Goetzmann (1995). “Performance persistence”. *The Journal of Finance* 50 (2), pp. 679–698.
Brown, S., M. Kang, F. In, and G. Lee (2010). “Resisting the manipulation of performance metrics: an empirical analysis of the manipulation-proof performance measure”. Working paper. URL: https://ssrn.com/abstract=1536323.
Burke, G. (1994). “A sharper Sharpe ratio”. *Futures* 23 (3), p. 56.
Capon, N., G. J. Fitzsimons, and R. A. Prince (1996). “An individual level analysis of the mutual fund investment decision”. *Journal of Financial Services Research* 10 (1), pp. 59–82.
Carhart, M. M. (1997). “On persistence in mutual fund performance”. *The Journal of Finance* 52 (1), pp. 57–82.
Carles, P. G., L. M. Doncel, and J. Sainz (2018). “Stability in mutual fund performance rankings: a new proposal”. *International Review of Economics & Finance*.
Ding, B. and H. A. Shawky (2007). “The performance of hedge fund strategies and the asymmetry of return distributions”. *European Financial Management* 13 (2), pp. 309–331.
Dittmar, R. F. (2002). “Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns”. *The Journal of Finance* 57 (1), pp. 369–403.
Dowd, K. (2000). “Adjusting for risk: An improved Sharpe ratio”. *International Review of Economics & Finance* 9 (3), pp. 209–222.
Eling, M. (2008). “Does the measure matter in the mutual fund industry?” *Financial Analysts Journal* 64 (3), pp. 54–66.
Eling, M., S. Farinelli, D. Rossello, and L. Tibiletti (2011). “One-size or tailor-made performance ratios for ranking hedge funds?” *Journal of Derivatives & Hedge Funds* 16 (4), pp. 267–277.
Eling, M. and R. Faust (2010). “The performance of hedge funds and mutual funds in emerging markets”. *Journal of Banking & Finance* 34 (8), pp. 1993–2009.
Eling, M. and F. Schuhmacher (2007). “Does the choice of performance measure influence the evaluation of hedge funds?” *Journal of Banking & Finance* 31 (9), pp. 2632–2647.
Elton, E. J., M. J. Gruber, and C. R. Blake (1996). “The persistence of risk-adjusted mutual fund performance”. *The Journal of Business* 69 (2), pp. 133–157.
— (2001). “A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases”. *The Journal of Finance* 56 (6), pp. 2415–2430.

Elton, E. J., M. J. Gruber, and C. R. Blake (2012). "Does mutual fund size matter? The relationship between size and performance". *The Review of Asset Pricing Studies* 2 (1), pp. 31–55.

Evans, R. B. (2010). "Mutual fund incubation". *The Journal of Finance* 65 (4), pp. 1581–1611.

Fama, E. F. and K. R. French (1993). "Common risk factors in the returns on stocks and bonds". *Journal of Financial Economics* 33 (1), pp. 3–56.

— (2015). "A five-factor asset pricing model". *Journal of Financial Economics* 116 (1), pp. 1–22.

Farinelli, S. and L. Tibiletti (2008). "Sharpe thinking in asset ranking with one-sided measures". *European Journal of Operational Research* 185 (3), pp. 1542–1547.

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

Guo, B. and Y. Xiao (2016). "A note on why doesn't the choice of performance measure matter?" *Finance Research Letters* 16, pp. 248–254.

Harvey, C. and A. Siddique (2000). "Conditional skewness in asset pricing tests". *The Journal of Finance* 55 (3), pp. 1263–1295.

Hendricks, D., J. Patel, and R. Zeckhauser (1993). "Hot hands in mutual funds: short-run persistence of relative performance, 1974–1988". *The Journal of Finance* 48 (1), pp. 93–130.

Kaplan, P. D. and J. A. Knowles (2004). "Kappa: a generalized downside risk-adjusted performance measure". *Journal of Performance Measurement* 8, pp. 42–54.

Keating, C. and W. F. Shadwick (2002). "A universal performance measure". *Journal of Performance Measurement* 6 (3), pp. 59–84.

Kestner, L. N. (1996). "Getting a handle on true performance". *Futures* 25, pp. 44–47.

Kostakis, A. (2009). "Performance measures and incentives: loading negative coskewness to outperform the CAPM". *The European Journal of Finance* 15 (5-6), pp. 463–486.

Kraus, A. and R. H. Litzenberger (1976). "Skewness preference and the valuation of risk assets". *The Journal of Finance* 31 (4), pp. 1085–1100.

Martin, R. D., S. Z. Rachev, and F. Siboulet (2003). "Phi-alpha optimal portfolios and extreme risk management". *Wilmott Magazine* November/December, pp. 70–83.

Moreno, D. and R. Rodríguez (2009). "The value of coskewness in mutual fund performance evaluation". *Journal of Banking & Finance* 33 (9), pp. 1664–1676.

Newey, W. K. and K. D. West (1987). "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix". *Econometrica* 55 (3), pp. 703–708.

— (1994). "Automatic lag selection in covariance matrix estimation". *The Review of Economic Studies* 61 (4), pp. 631–653.

Ornelas, J. R. H., A. F. Silva Júnior, and J. L. B. Fernandes (2012). "Yes, the choice of performance measure does matter for ranking of us mutual funds". *International Journal of Finance & Economics* 17 (1), pp. 61–72.

Pedersen, C. S. and S. E. Satchell (2002). “On the foundation of performance measures under asymmetric returns”. *Quantitative Finance* 2 (3), pp. 217–223.
Rachev, S., T. Jašić, S. Stoyanov, and F. J. Fabozzi (2007). “Momentum strategies based on reward–risk stock selection criteria”. *Journal of Banking & Finance* 31 (8), pp. 2325–2346.
Schuhmacher, F. and M. Eling (2012). “A decision-theoretic foundation for reward-to-risk performance measures”. *Journal of Banking & Finance* 36 (7), pp. 2077–2082.
Sharpe, W. F. (1966). “Mutual fund performance”. *Journal of Business*, pp. 119–138.
Sortino, F. A., R. v. d. Meer, and A. Plantinga (1999). “The dutch triangle”. *The Journal of Portfolio Management* 26 (1), pp. 50–57.
Treynor, J. L. and F. Black (1973). “How to use security analysis to improve portfolio selection”. *The Journal of Business*, pp. 66–86.
Young, T. (1991). “Calmar ratio: a smoother tool”. *Futures* 20 (1), p. 40.
Zakamouline, V. (2011). “The performance measure you choose influences the evaluation of hedge funds”. *Journal of Performance Measurement* 15 (3), pp. 48–64.

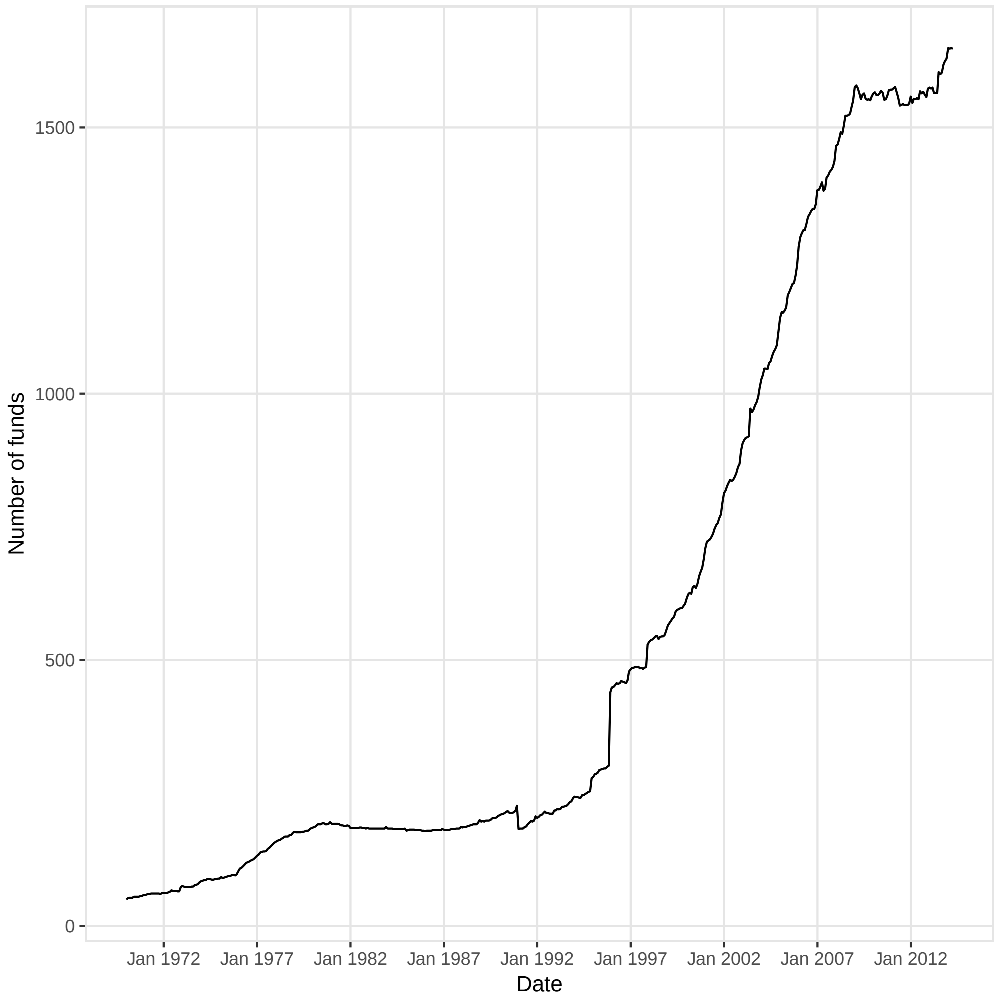

Figure 1 – Number of funds per month over the period from December 1969 to March 2014

Table 1 – Average excess returns over the market benchmark

	1	2	3	4	5	6	7	8	9	10
Sharpe	**0.6187**	**0.7147**	**0.7252**	**0.7477**	**0.7623**	**0.8046**	**0.8367**	**0.8536**	**0.8813**	**0.9318**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
ERVaR-HS(0.05)	**0.6290**	**0.7143**	**0.7357**	**0.7304**	**0.7771**	**0.8150**	**0.8057**	**0.8552**	**0.8936**	**0.9186**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
ERVaR-NG(0.05)	**0.6267**	**0.7081**	**0.7125**	**0.7525**	**0.7373**	**0.8235**	**0.8348**	**0.8558**	**0.9018**	**0.9176**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
ERCVaR-HS(0.05)	**0.6173**	**0.7395**	**0.7222**	**0.7586**	**0.7765**	**0.7883**	**0.8334**	**0.8626**	**0.8455**	**0.9270**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
ModSortino	**0.6254**	**0.7103**	**0.7342**	**0.7480**	**0.7627**	**0.8287**	**0.8254**	**0.8607**	**0.8548**	**0.9265**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
FT(1,1)	**0.6276**	**0.6964**	**0.7396**	**0.7562**	**0.7417**	**0.8167**	**0.8355**	**0.8711**	**0.8591**	**0.9300**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
FT(1,2)	**0.6393**	**0.6850**	**0.7721**	**0.7695**	**0.7240**	**0.8234**	**0.8357**	**0.8615**	**0.8567**	**0.9055**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
RR(0.05,0.05)	**0.7691**	**0.7724**	**0.8096**	**0.7566**	**0.7762**	**0.7953**	**0.7933**	**0.8279**	**0.7975**	**0.7713**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
Calmar	**0.6157**	**0.7150**	**0.7699**	**0.7566**	**0.7795**	**0.8158**	**0.8310**	**0.8878**	**0.8174**	**0.8869**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
Sterling	**0.6200**	**0.7375**	**0.7128**	**0.7575**	**0.7716**	**0.8058**	**0.8137**	**0.8612**	**0.8683**	**0.9231**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
Burke	**0.6147**	**0.7402**	**0.7183**	**0.7658**	**0.7679**	**0.8066**	**0.8325**	**0.8538**	**0.8598**	**0.9142**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
AR-CAPM	**0.6101**	**0.6698**	**0.7058**	**0.7659**	**0.7782**	**0.8351**	**0.8631**	**0.8278**	**0.9091**	**0.9064**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
MPPM(3)	**0.7041**	**0.7379**	**0.7211**	**0.7820**	**0.7756**	**0.7856**	**0.8240**	**0.8416**	**0.8407**	**0.8693**
	(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**

This Table reports mean excess returns over the market benchmark and the corresponding p-values of the paired t-test of the mean being equal to zero for equally weighted decile portfolios of funds ranked according to the different risk-adjusted performance measures. The t-test is performed with heteroscedasticity and autocorrelation consistent standard errors following Newey and West (1987, 1994). The market benchmark is US market return available on the Professor Kenneth French webpage. Portfolios are rebalanced monthly. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model. Decile 1 corresponds to the bottom performing funds and decile 10 to the top performing funds. Both mean excess returns and p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.

Table 2 – Four-factor alphas for equally weighted portfolios

	1	2	3	4	5	6	7	8	9	10
Sharpe	**-0.1671**	**-0.1217**	**-0.1053**	-0.0958	**-0.0854**	-0.0708	-0.0614	-0.0358	-0.0155	0.0467
	(3.11)	**(3.90)**	**(4.00)**	(5.65)	**(3.38)**	(12.12)	(17.85)	(44.12)	(76.00)	(34.82)
ERVaR-HS(0.05)	**-0.1574**	**-0.1128**	**-0.1097**	**-0.0990**	-0.0762	-0.0550	-0.0921	-0.0219	-0.0121	0.0213
	(4.52)	**(4.72)**	**(3.15)**	**(3.56)**	(8.07)	(18.05)	(5.07)	(61.75)	(81.56)	(67.70)
ERVaR-NG(0.05)	**-0.1636**	**-0.1320**	**-0.1236**	**-0.0912**	**-0.0987**	-0.0616	-0.0674	-0.0268	0.0099	0.0380
	(1.79)	**(1.78)**	**(1.27)**	**(3.52)**	**(1.39)**	(21.74)	(13.05)	(54.23)	(84.02)	(42.75)
ERCVaR-HS(0.05)	**-0.1616**	-0.1051	**-0.1138**	-0.0775	-0.0654	**-0.0879**	-0.0495	-0.0431	-0.0532	0.0384
	(3.89)	(7.20)	**(3.39)**	(11.63)	(13.74)	**(3.70)**	(26.82)	(31.51)	(26.75)	(50.31)
ModSortino	**-0.1561**	**-0.1284**	**-0.1009**	-0.0863	**-0.0844**	-0.0496	-0.0668	-0.0318	-0.0481	0.0398
	(4.60)	**(3.62)**	**(4.53)**	(8.88)	**(4.02)**	(30.21)	(13.38)	(46.63)	(32.11)	(46.26)
FT(1,1)	**-0.1539**	**-0.1355**	-0.0926	-0.0818	**-0.1130**	-0.0535	-0.0674	-0.0170	-0.0361	0.0369
	(4.90)	**(3.02)**	(7.53)	(12.02)	**(0.44)**	(26.39)	(13.94)	(70.98)	(47.60)	(48.07)
FT(1,2)	**-0.1623**	**-0.1434**	-0.0598	-0.0778	**-0.1133**	-0.0407	-0.0486	-0.0285	-0.0440	0.0052
	(3.38)	**(1.62)**	(25.18)	(11.37)	**(0.60)**	(38.90)	(26.94)	(52.36)	(32.03)	(92.95)
RR(0.05,0.05)	-0.0815	-0.0820	-0.0375	-0.0791	-0.0625	-0.0643	-0.0639	-0.0574	**-0.1083**	-0.0835
	(20.13)	(19.63)	(50.22)	(9.28)	(21.47)	(18.45)	(21.01)	(26.66)	**(1.82)**	(19.60)
Calmar	**-0.1666**	**-0.1301**	-0.0771	-0.0883	-0.0658	-0.0605	-0.0679	0.0020	-0.0696	0.0098
	(2.59)	**(3.43)**	(15.41)	(7.41)	(14.65)	(19.33)	(9.90)	(96.47)	(12.91)	(86.85)
Sterling	**-0.1587**	-0.1082	**-0.1242**	-0.0797	-0.0666	-0.0744	-0.0786	-0.0245	-0.0377	0.0361
	(4.11)	(7.17)	**(0.97)**	(13.34)	(11.07)	(8.78)	(7.38)	(56.99)	(42.23)	(49.99)
Burke	**-0.1636**	-0.1081	**-0.1172**	-0.0691	-0.0780	-0.0767	-0.0495	-0.0405	-0.0454	0.0314
	(3.32)	(7.29)	**(2.64)**	(18.21)	(7.88)	(8.65)	(27.73)	(31.97)	(34.82)	(57.99)
AR-CAPM	**-0.1716**	**-0.1262**	**-0.1173**	-0.0758	-0.0799	-0.0669	-0.0311	-0.0854	0.0147	0.0211
	(0.48)	**(3.25)**	**(2.31)**	(17.72)	(11.34)	(17.91)	(54.34)	(7.46)	(76.95)	(66.38)
MPPM(3)	-0.1336	-0.0967	**-0.1257**	-0.0631	-0.0504	-0.0678	-0.0577	-0.0466	-0.0349	-0.0277
	(10.39)	(13.88)	**(2.31)**	(19.54)	(24.95)	(9.19)	(21.92)	(25.30)	(41.54)	(60.46)

This Table reports the Carhart (1997) four-factor alphas and the corresponding p-values of the t-test of alpha being equal to zero for equally weighted decile portfolios of funds ranked according to the different risk-adjusted performance measures. The reported p-values are based on standard errors corrected for the presence of autocorrelation and heteroscedasticity using the procedure suggested by Newey and West (1987, 1994). Portfolios are rebalanced monthly. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model. Decile 1 corresponds to the bottom performing funds and decile 10 to the top performing funds. Both alphas and p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.

Table 3 – Five-factor alphas for equally weighted portfolios

	1	2	3	4	5	6	7	8	9	10
Sharpe	**-0.2150**	**-0.1328**	**-0.1249**	**-0.1334**	**-0.1013**	-0.0645	-0.0686	-0.0415	-0.0142	0.0666
	(0.13)	**(1.34)**	**(0.29)**	**(0.31)**	**(0.97)**	(15.80)	(15.09)	(39.07)	(79.51)	(22.90)
ERVaR-HS(0.05)	**-0.2060**	**-0.1169**	**-0.1321**	**-0.1218**	-0.0727	-0.0777	**-0.1031**	-0.0383	-0.0230	0.0580
	(0.07)	**(1.76)**	**(0.20)**	**(0.63)**	(9.15)	(6.97)	**(3.52)**	(43.64)	(67.23)	(29.59)
ERVaR-NG(0.05)	**-0.1644**	-0.1080	**-0.1258**	**-0.1191**	**-0.1281**	-0.0702	-0.0847	-0.0517	-0.0081	0.0254
	(1.69)	(5.79)	**(0.65)**	**(0.29)**	**(0.09)**	(13.82)	(6.76)	(29.24)	(88.09)	(64.75)
ERCVaR-HS(0.05)	**-0.2336**	**-0.1064**	**-0.1388**	**-0.1251**	**-0.0845**	**-0.0935**	-0.0570	-0.0371	-0.0575	0.0962
	(0.04)	**(3.12)**	**(0.15)**	**(0.57)**	**(4.47)**	**(2.36)**	(20.84)	(45.43)	(26.85)	(11.67)
ModSortino	**-0.2156**	**-0.1408**	**-0.1266**	**-0.1433**	**-0.0997**	-0.0592	-0.0673	-0.0300	-0.0440	0.0965
	(0.10)	**(0.73)**	**(0.24)**	**(0.14)**	**(0.82)**	(22.47)	(15.78)	(54.63)	(41.75)	(8.26)
FT(1,1)	**-0.2054**	**-0.1485**	**-0.1160**	**-0.1236**	**-0.1406**	-0.0560	-0.0666	-0.0172	-0.0341	0.0762
	(0.17)	**(0.65)**	**(0.61)**	**(0.71)**	**(0.05)**	(23.81)	(15.61)	(70.59)	(52.84)	(16.59)
FT(1,2)	**-0.2277**	**-0.1775**	**-0.0898**	**-0.1197**	**-0.1506**	-0.0450	-0.0547	-0.0139	-0.0294	0.0763
	(0.03)	**(0.08)**	**(4.23)**	**(0.88)**	**(0.03)**	(31.92)	(21.71)	(77.50)	(57.40)	(20.39)
RR(0.05,0.05)	-0.1066	**-0.1175**	**-0.0908**	**-0.1414**	**-0.1110**	**-0.1109**	**-0.1044**	-0.0682	-0.0546	0.0694
	(5.27)	**(2.75)**	**(4.84)**	**(0.20)**	**(1.86)**	**(1.07)**	**(1.70)**	(13.48)	(19.24)	(20.88)
Calmar	**-0.2397**	**-0.1577**	**-0.1034**	**-0.1394**	**-0.0898**	-0.0749	-0.0684	0.0073	-0.0505	0.0860
	(0.03)	**(0.18)**	**(2.18)**	**(0.28)**	**(2.31)**	(10.54)	(14.40)	(87.36)	(31.42)	(17.47)
Sterling	**-0.2187**	**-0.1122**	**-0.1407**	**-0.1186**	-0.0799	**-0.0883**	-0.0774	-0.0361	-0.0379	0.0745
	(0.11)	**(2.74)**	**(0.09)**	**(1.46)**	(5.08)	**(4.09)**	(8.48)	(47.09)	(48.06)	(18.90)
Burke	**-0.2294**	**-0.1174**	**-0.1382**	**-0.1175**	**-0.0935**	**-0.0859**	-0.0534	-0.0401	-0.0445	0.0857
	(0.05)	**(2.32)**	**(0.12)**	**(1.18)**	**(3.67)**	**(3.91)**	(25.77)	(38.08)	(40.06)	(15.59)
AR-CAPM	**-0.2189**	**-0.1646**	**-0.1481**	**-0.1201**	**-0.0953**	-0.0571	-0.0224	-0.0797	0.0322	0.0411
	(0.01)	**(0.14)**	**(0.05)**	**(1.29)**	**(2.85)**	(22.61)	(64.67)	(11.43)	(54.44)	(42.61)
MPPM(3)	**-0.1558**	**-0.1654**	**-0.1850**	**-0.1445**	**-0.1089**	**-0.0890**	-0.0703	-0.0252	0.0080	0.1149
	(2.51)	**(0.95)**	**(0.02)**	**(0.19)**	**(1.04)**	**(1.62)**	(16.17)	(57.56)	(86.45)	(5.50)

This Table reports the Fama and French (2015) five-factor alphas and the corresponding p-values of the t-test of alpha being equal to zero for equally weighted decile portfolios of funds ranked according to the different risk-adjusted performance measures. The reported p-values are based on standard errors corrected for the presence of autocorrelation and heteroscedasticity using the procedure suggested by Newey and West (1987, 1994). Portfolios are rebalanced monthly. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model. Decile 1 corresponds to the bottom performing funds and decile 10 to the top performing funds. Both alphas and p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.

Table 4 – Differences in *ex post* performance between the top (D10) and the bottom (D1) performing funds: monthly rebalancing

	Equally weighted portfolios						Value weighted portfolios					
	D10-D1 Mean excess returns		D10-D1 Fama and French (2015) five-factor alphas		D10-D1 Carhart (1997) four-factor alphas		D10-D1 Mean excess returns		D10-D1 Fama and French (2015) five-factor alphas		D10-D1 Carhart (1997) four-factor alphas	
Sharpe	**0.3131**	**(0.04)**	**0.2816**	**(0.30)**	**0.2138**	**(2.68)**	**0.2499**	**(0.63)**	**0.2531**	**(0.67)**	0.1687	7.99
ERVaR-HS(0.05)	**0.2896**	**(0.08)**	**0.2639**	**(0.49)**	0.1787	(6.84)	**0.2449**	**(0.8)**	**0.2413**	**(1.17)**	0.1481	15.39
ERVaR-NG(0.05)	**0.2909**	**(0.15)**	0.1897	(5.21)	**0.2016**	**(1.98)**	**0.2817**	**(0.19)**	**0.2020**	**(4.41)**	**0.2036**	**(1.82)**
ERCVaR-HS(0.05)	**0.3097**	**(0.09)**	**0.3297**	**(0.09)**	0.2000	(5.23)	**0.2678**	**(0.57)**	**0.3210**	**(0.08)**	0.1716	9.89
ModSortino	**0.3011**	**(0.10)**	**0.3121**	**(0.13)**	0.1959	(5.77)	**0.2311**	**(1.95)**	**0.2725**	**(0.41)**	0.1452	17.57
FT(1,1)	**0.3023**	**(0.09)**	**0.2817**	**(0.35)**	0.1908	(6.62)	**0.2195**	**(1.62)**	**0.2340**	**(0.91)**	0.1241	21.74
FT(1,2)	**0.2662**	**(0.35)**	**0.3040**	**(0.12)**	0.1675	1(0.10)	**0.2399**	**(1.24)**	**0.3001**	**(0.09)**	0.1461	15.5
RR(0.05,0.05)	0.0023	(97.39)	**0.1760**	**(2.34)**	-0.0020	9(8.28)	-0.0075	(92.88)	0.1440	(9.47)	-0.0298	75.14
Calmar	**0.2712**	**(0.49)**	**0.3257**	**(0.14)**	0.1764	(9.34)	**0.2293**	**(2.19)**	**0.3430**	**(0.07)**	0.1532	15.64
Sterling	**0.3031**	**(0.08)**	**0.2932**	**(0.19)**	0.1948	(5.14)	**0.2421**	**(1.26)**	**0.2805**	**(0.47)**	0.1579	14.49
Burke	**0.2995**	**(0.11)**	**0.3151**	**(0.15)**	0.1950	(5.97)	**0.2406**	**(1.6)**	**0.2944**	**(0.31)**	0.1635	12.98
AR-CAPM	**0.2963**	**(0.01)**	**0.2600**	**(0.11)**	**0.1927**	**(2.24)**	**0.2245**	**(0.07)**	**0.2213**	**(0.14)**	**0.1536**	**(3.77)**
MPPM(3)	0.1652	(11.48)	**0.2707**	**(1.48)**	0.1059	2(9.29)	0.1472	(18.34)	0.2108	(5.02)	0.0883	(42.39)

This Table reports the *ex post* performance of the differences portfolios between the top performing funds (D10) and the bottom performing funds (D1) considering the different risk-adjusted performance measures. Portfolios are rebalanced monthly. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model. *Ex post* performance is measured by excess returns, Fama and French (2015) five-factor alphas and Carhart (1997) four-factor alphas. Excess returns, alphas and their corresponding p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.

Table 5 – Mean excess returns of the differences portfolios between pairs of performance measures for bottom performing funds (D1)

	ERVaR-HS(0.05)	ERVaR-NG(0.05)	ERCVaR-HS(0.05)	ModSortino	FT(1,1)	FT(1,2)	RR(0.05,0.05)	Calmar	Sterling	Burke	AR-CAPM	MPPM(3)
Sharpe	0.0103	0.0080	-0.0014	0.0067	0.0090	0.0207	**0.1504**	-0.0030	0.0014	-0.0040	-0.0086	0.0854
	(42.90)	(66.51)	(90.66)	(45.13)	(26.48)	(44.38)	**(0.27)**	(87.15)	(89.79)	(73.44)	(82.51)	(6.64)
ERVaR-HS (0.05)		-0.0023	-0.0117	-0.0036	-0.0013	0.0103	**0.1401**	-0.0133	-0.0089	-0.0143	-0.0189	0.0751
	-	(90.66)	(41.28)	(81.68)	(92.43)	(71.00)	**(0.56)**	(57.47)	(51.98)	(33.15)	(65.59)	(11.31)
ERVaR-NG (0.05)			-0.0093	-0.0013	0.0010	0.0127	**0.1424**	-0.0109	-0.0066	-0.0120	-0.0166	0.0774
	-	-	(62.56)	(94.69)	(94.29)	(70.85)	**(0.82)**	(71.24)	(68.14)	(52.80)	(69.34)	(11.59)
ERCVaR-HS (0.05)				0.0081	0.0103	0.0220	**0.1518**	-0.0016	0.0027	-0.0026	-0.0072	0.0867
	-	-	-	(39.25)	(41.16)	(44.42)	**(0.27)**	(90.90)	(54.85)	(52.92)	(85.94)	(8.24)
ModSortino					0.0023	0.0139	**0.1437**	-0.0097	-0.0053	-0.0107	-0.0153	0.0787
	-	-	-	-	(83.08)	(61.99)	**(0.46)**	(48.96)	(50.21)	(26.68)	(70.03)	(10.43)
FT (1,1)						0.0117	**0.1414**	-0.0119	-0.0076	-0.0129	-0.0176	0.0764
	-	-	-	-	-	(65.98)	**(0.42)**	(52.56)	(47.98)	(33.51)	(64.89)	(10.06)
FT (1,2)							**0.1297**	-0.0236	-0.0193	-0.0246	-0.0292	0.0647
	-	-	-	-	-	-	**(0.32)**	(42.67)	(50.11)	(38.80)	(51.12)	(22.02)
RR (0.05,0.05)								**-0.1534**	**-0.1490**	**-0.1544**	**-0.1590**	-0.0650
	-	-	-	-	-	-	-	**(0.26)**	**(0.35)**	**(0.22)**	**(0.08)**	(30.81)
Calmar									0.0043	-0.0010	-0.0056	0.0883
	-	-	-	-	-	-	-	-	(75.96)	(94.43)	(88.89)	(9.39)
Sterling										-0.0053	-0.0100	0.0840
	-	-	-	-	-	-	-	-	-	(32.31)	(80.35)	(8.93)
Burke											-0.0046	0.0894
	-	-	-	-	-	-	-	-	-	-	(90.96)	(7.23)
AR-CAPM												0.0940
	-	-	-	-	-	-	-	-	-	-	-	(9.40)

This Table reports the mean excess returns and the corresponding p-values of the paired t-test of the mean being equal to zero for the differences portfolios between each pair of performance measures (measure in column minus measure in line) considering the bottom performing funds (equally weighted decile 1). The t-test is performed with heteroscedasticity and autocorrelation consistent standard errors following Newey and West (1987, 1994). Portfolios are rebalanced monthly. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model. Both alphas and p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.

Table 6 – Mean excess returns of the differences portfolios between pairs of performance measures for top performing funds (D10)

	ERVaR-HS(0.05)	ERVaR-NG(0.05)	ERCVaR-HS(0.05)	ModSortino	FT(1,1)	FT(1,2)	RR(0.05,0.05)	Calmar	Sterling	Burke	AR-CAPM	MPPM(3)
Sharpe	-0.0132	-0.0142	-0.0047	-0.0053	-0.0018	-0.0263	**-0.1605**	-0.0448	-0.0086	-0.0175	-0.0254	-0.0625
	(28.16)	(49.21)	(78.76)	(69.06)	(85.78)	(47.05)	**(1.62)**	(9.82)	(49.39)	(28.75)	(19.19)	(13.67)
ERVaR-HS (0.05)		-0.0010	0.0085	0.0079	0.0114	-0.0131	**-0.1473**	-0.0316	0.0046	-0.0043	-0.0122	-0.0493
	-	(96.64)	(64.84)	(60.22)	(46.56)	(68.17)	**(1.94)**	(24.83)	(74.32)	(81.44)	(57.59)	(21.23)
ERVaR-NG (0.05)			0.0095	0.0090	0.0124	-0.0120	**-0.1462**	-0.0306	0.0056	-0.0033	-0.0112	-0.0483
	-	-	(76.06)	(73.48)	(47.79)	(77.38)	**(2.77)**	(39.71)	(82.54)	(90.47)	(66.17)	(29.49)
ERCVaR-HS (0.05)				-0.0005	0.0029	-0.0215	**-0.1557**	**-0.0401**	-0.0039	-0.0128	-0.0207	-0.0577
	-	-	-	(96.17)	(89.27)	(48.17)	**(1.02)**	**(1.89)**	(76.69)	(9.54)	(42.36)	(10.61)
ModSortino					0.0034	-0.0210	**-0.1552**	-0.0396	-0.0034	-0.0123	-0.0202	-0.0572
	-	-	-	-	(82.62)	(48.86)	**(0.89)**	(6.23)	(72.57)	(15.58)	(33.28)	(10.33)
FT (1,1)						-0.0244	**-0.1586**	-0.0430	-0.0068	-0.0157	-0.0236	-0.0607
	-	-	-	-	-	(48.79)	**(1.14)**	(16.48)	(67.14)	(40.45)	(19.98)	(12.65)
FT (1,2)							**-0.1342**	-0.0186	0.0176	0.0087	0.0008	-0.0362
	-	-	-	-	-	-	**(0.49)**	(52.55)	(60.21)	(75.92)	(98.11)	(37.69)
RR (0.05,0.05)								**0.1156**	**0.1518**	**0.1429**	**0.1350**	0.0980
	-	-	-	-	-	-	-	**(4.12)**	**(1.02)**	**(1.45)**	**(2.48)**	(9.29)
Calmar									0.0362	0.0273	0.0194	-0.0176
	-	-	-	-	-	-	-	-	(13.36)	(8.87)	(52.53)	(62.56)
Sterling										-0.0089	-0.0168	-0.0539
	-	-	-	-	-	-	-	-	-	(44.54)	(41.61)	(17.09)
Burke											-0.0079	-0.0450
	-	-	-	-	-	-	-	-	-	-	(73.53)	(20.64)
AR-CAPM												-0.0371
	-	-	-	-	-	-	-	-	-	-	-	(37.52)

This Table reports the mean excess returns and the corresponding p-values of the paired t-test of the mean being equal to zero for the differences portfolios between each pair of performance measures (measure in column minus measure in line) considering the top performing funds (equally weighted decile 10). The t-test is performed with heteroscedasticity and autocorrelation consistent standard errors following Newey and West (1987, 1994). Portfolios are rebalanced monthly. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model. Both alphas and p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.

Table 7 – Carhart (1997) four-factor alphas of the differences portfolios between pairs of performance measures for bottom performing funds (D1)

	ERVaR-HS(0.05)	ERVaR-NG(0.05)	ERCVaR-HS(0.05)	ModSortino	FT(1,1)	FT(1,2)	RR(0.05,0.05)	Calmar	Sterling	Burke	AR-CAPM	MPPM(3)
Sharpe	0.0097	0.0035	0.0055	0.0110	0.0132	0.0048	0.0856	0.0005	0.0084	0.0035	-0.0045	0.0335
	(41.80)	(85.34)	(65.36)	(12.65)	(10.65)	(83.87)	(9.33)	(97.72)	(39.19)	(76.00)	(89.34)	(42.29)
ERVaR-HS (0.05)		-0.0062	-0.0042	0.0013	0.0035	-0.0049	0.0759	-0.0092	-0.0013	-0.0062	-0.0142	0.0238
	-	(77.45)	(71.12)	(92.32)	(77.67)	(84.85)	(15.22)	(66.01)	(90.93)	(60.94)	(70.75)	(57.16)
ERVaR-NG (0.05)			0.0020	0.0074	0.0097	0.0013	0.0820	-0.0030	0.0048	-0.0000	-0.0080	0.0300
	-	-	(92.43)	(71.28)	(59.85)	(96.14)	(9.84)	(88.59)	(79.84)	(99.88)	(80.87)	(46.39)
ERCVaR-HS (0.05)				0.0055	0.0077	-0.0007	0.0801	-0.0050	0.0029	-0.0020	-0.0100	0.0280
	-	-	-	(56.71)	(57.83)	(97.89)	(12.22)	(74.72)	(68.38)	(76.53)	(76.98)	(50.86)
ModSortino					0.0022	-0.0062	0.0746	-0.0105	-0.0026	-0.0075	-0.0155	0.0226
	-	-	-	-	(81.65)	(80.07)	(14.52)	(50.75)	(74.98)	(42.14)	(65.37)	(59.61)
FT (1,1)						-0.0084	0.0724	-0.0127	-0.0048	-0.0097	-0.0177	0.0203
	-	-	-	-	-	(71.79)	(14.71)	(49.03)	(66.90)	(48.29)	(60.34)	(63.62)
FT (1,2)							0.0808	-0.0043	0.0036	-0.0013	-0.0093	0.0287
	-	-	-	-	-	-	(6.18)	(87.45)	(89.01)	(95.98)	(80.49)	(54.83)
RR (0.05,0.05)								-0.0851	-0.0772	-0.0821	**-0.0901**	-0.0521
	-	-	-	-	-	-	-	(9.71)	(13.36)	(10.76)	**(4.51)**	(41.43)
Calmar									0.0079	0.0030	-0.0050	0.0330
	-	-	-	-	-	-	-	-	(61.37)	(85.28)	(88.07)	(44.45)
Sterling										-0.0049	-0.0129	0.0251
	-	-	-	-	-	-	-	-	-	(41.45)	(70.82)	(55.64)
Burke											-0.0080	0.0300
	-	-	-	-	-	-	-	-	-	-	(81.08)	(48.81)
AR-CAPM												0.0380
	-	-	-	-	-	-	-	-	-	-	-	(42.70)

This Table reports the Carhart (1997) four-factor alphas and the corresponding p-values of the t-test of the alpha being equal to zero for the differences portfolios between each pair of performance measures (measure in column minus measure in line) considering the bottom performing funds (equally weighted decile 1). The reported p-values are based on standard errors corrected for the presence of autocorrelation and heteroscedasticity using the procedure suggested by Newey and West (1987, 1994). Portfolios are rebalanced monthly. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model. Both alphas and p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.

Table 8 – Carhart (1997) four-factor alphas of the differences portfolios between pairs of performance measures for top performing funds (D10)

	ERVaR-HS(0.05)	ERVaR-NG(0.05)	ERCVaR-HS(0.05)	ModSortino	FT(1,1)	FT(1,2)	RR(0.05,0.05)	Calmar	Sterling	Burke	AR-CAPM	MPPM(3)
Sharpe	-0.0254	-0.0087	-0.0083	-0.0069	-0.0097	-0.0415	**-0.1302**	-0.0369	-0.0106	-0.0153	-0.0256	**-0.0744**
	(9.08)	(67.53)	(64.26)	(65.22)	(36.15)	(14.47)	**(3.92)**	(15.59)	(48.54)	(39.89)	(22.16)	**(4.39)**
ERVaR-HS (0.05)		0.0167	0.0171	0.0185	0.0157	-0.0161	-0.1048	-0.0114	0.0148	0.0102	-0.0002	-0.0490
	-	(49.61)	(41.96)	(38.34)	(35.41)	(61.31)	(9.30)	(70.24)	(35.35)	(66.46)	(99.40)	(20.39)
ERVaR-NG (0.05)			0.0004	0.0018	-0.0011	-0.0328	-0.1215	-0.0282	-0.0019	-0.0066	-0.0169	-0.0657
	-	-	(99.14)	(95.21)	(96.58)	(37.46)	(6.14)	(42.76)	(94.65)	(83.33)	(52.22)	(11.93)
ERCVaR-HS (0.05)				0.0014	-0.0014	-0.0332	**-0.1219**	-0.0286	-0.0023	-0.0070	-0.0173	-0.0661
	-	-	-	(88.64)	(94.21)	(18.41)	**(3.57)**	(5.85)	(84.47)	(44.43)	(46.42)	(6.43)
ModSortino					-0.0028	-0.0346	**-0.1233**	-0.0299	-0.0037	-0.0083	-0.0187	-0.0675
	-	-	-	-	(85.99)	(14.63)	**(2.98)**	(6.81)	(73.59)	(31.75)	(36.64)	(5.61)
FT (1,1)						-0.0318	**-0.1205**	-0.0271	-0.0009	-0.0055	-0.0158	-0.0647
	-	-	-	-	-	(24.52)	**(4.61)**	(29.28)	(95.95)	(75.80)	(31.63)	(6.59)
FT (1,2)							-0.0887	0.0046	0.0309	0.0262	0.0159	-0.0329
	-	-	-	-	-	-	(6.68)	(84.67)	(22.84)	(22.00)	(60.67)	(46.18)
RR (0.05,0.05)								0.0933	**0.1196**	**0.1149**	0.1046	0.0558
	-	-	-	-	-	-	-	(9.59)	**(4.10)**	**(3.82)**	(7.11)	(33.50)
Calmar									0.0262	0.0216	0.0113	-0.0375
	-	-	-	-	-	-	-	-	(19.35)	(8.13)	(66.73)	(31.38)
Sterling										-0.0046	-0.0150	-0.0638
	-	-	-	-	-	-	-	-	-	(71.35)	(45.55)	(8.44)
Burke											-0.0103	-0.0591
	-	-	-	-	-	-	-	-	-	-	(64.64)	(10.51)
AR-CAPM												-0.0488
	-	-	-	-	-	-	-	-	-	-	-	(19.22)

This Table reports the Carhart (1997) four-factor alphas and the corresponding p-values of the t-test of the alpha being equal to zero for the differences portfolios between each pair of performance measures (measure in column minus measure in line) considering the top performing funds (equally weighted decile 10). The reported p-values are based on standard errors corrected for the presence of autocorrelation and heteroscedasticity using the procedure suggested by Newey and West (1987, 1994). Portfolios are rebalanced monthly. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model. Both alphas and p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.

Table 9 – Turnover for equally weighted portfolios: monthly rebalancing

	1	2	3	4	5	6	7	8	9	10
Sharpe	0.0919	0.1178	0.1344	0.1426	0.1425	0.1441	0.1374	0.1303	0.1086	0.0809
ERVaR-HS(0.05)	0.0951	0.1186	0.1380	0.1431	0.1423	0.1442	0.1383	0.1320	0.1114	0.0839
ERVaR-NG(0.05)	0.0990	0.1252	0.1408	0.1485	0.1466	0.1491	0.1428	0.1354	0.1172	0.0915
ERCVaR-HS(0.05)	0.0924	0.1180	0.1349	0.1430	0.1416	0.1444	0.1407	0.1291	0.1124	0.0834
ModSortino	0.0929	0.1166	0.1345	0.1429	0.1422	0.1439	0.1383	0.1299	0.1107	0.0812
FT(1,1)	0.0915	0.1176	0.1355	0.1415	0.1413	0.1435	0.1368	0.1284	0.1107	0.0805
FT(1,2)	0.0934	0.1179	0.1341	0.1419	0.1442	0.1451	0.1418	0.1308	0.1106	0.0842
RR(0.05,0.05)	0.0911	0.1180	0.1329	0.1347	0.1357	0.1342	0.1345	0.1268	0.1110	0.0866
Calmar	0.0933	0.1193	0.1363	0.1456	0.1434	0.1468	0.1391	0.1314	0.1129	0.0844
Sterling	0.0930	0.1174	0.1344	0.1407	0.1425	0.1438	0.1384	0.1299	0.1099	0.0826
Burke	0.0927	0.1177	0.1347	0.1421	0.1432	0.1433	0.1388	0.1294	0.1103	0.0835
AR-CAPM	0.0858	0.1103	0.1335	0.1429	0.1442	0.1489	0.1409	0.1303	0.1093	0.0809
MPPM(3)	0.0884	0.1141	0.1294	0.1383	0.1387	0.1428	0.1381	0.1268	0.1074	0.0799

This Table reports the average turnover for different deciles of equally weighted portfolios. Portfolios are rebalanced monthly. Decile 1 corresponds to the bottom performing funds and decile 10 to the top performing funds.

Table 10 – Differences in *ex post* performance between the top (D10) and the bottom (D1) performing funds: monthly rebalancing with transaction costs of 0.25% and 2% for equally weighted portfolios

	Transaction costs of 0.25%						Transaction costs of 2%					
	D10-D1 Mean excess returns		D10-D1 Fama and French (2015) five-factor alphas		D10-D1 Carhart (1997) four-factor alphas		D10-D1 Mean excess returns		D10-D1 Fama and French (2015) five-factor alphas		D10-D1 Carhart (1997) four-factor alphas	
Sharpe	**0.2699**	**(0.24)**	**0.2382**	**(1.20)**	0.1704	(7.78)	-0.0325	(71.93)	-0.0660	(49.09)	-0.1331	(17.93)
ERVaR-HS(0.05)	**0.2449**	**(0.47)**	**0.2189**	**(1.97)**	0.1339	(17.27)	-0.0684	(44.53)	-0.0960	(32.05)	-0.1795	(7.47)
ERVaR-NG(0.05)	**0.2433**	**(0.81)**	0.1421	(14.61)	0.1541	(7.50)	-0.0900	(32.86)	-0.1915	(5.59)	**-0.1786**	**(4.50)**
ERCVaR-HS(0.05)	**0.2658**	**(0.42)**	**0.2858**	**(0.38)**	0.1561	(12.99)	-0.0419	(66.39)	-0.0221	(82.55)	-0.1507	(14.95)
ModSortino	**0.2576**	**(0.49)**	**0.2684**	**(0.55)**	0.1523	(14.13)	-0.0470	(61.91)	-0.0376	(69.84)	-0.1529	(14.24)
FT(1,1)	**0.2593**	**(0.45)**	**0.2385**	**(1.34)**	0.1479	(15.49)	-0.0416	(65.36)	-0.0635	(51.46)	-0.1528	(14.72)
FT(1,2)	**0.2218**	**(1.50)**	**0.2599**	**(0.54)**	0.1231	(22.77)	-0.0889	(33.91)	-0.0491	(60.03)	-0.1872	(6.85)
RR(0.05,0.05)	-0.0422	(54.27)	0.1317	(8.95)	-0.0464	(61.55)	**-0.3531**	**(0.00)**	**-0.1779**	**(2.71)**	**-0.3576**	**(0.02)**
Calmar	**0.2268**	**(1.86)**	**0.2812**	**(0.58)**	0.1318	(21.01)	-0.0843	(39.26)	-0.0307	(76.42)	-0.1802	(9.08)
Sterling	**0.2592**	**(0.40)**	**0.2494**	**(0.82)**	0.1512	(13.09)	-0.0480	(60.06)	-0.0572	(55.02)	-0.1543	(13.22)
Burke	**0.2555**	**(0.54)**	**0.2713**	**(0.63)**	0.1510	(14.52)	-0.0529	(57.81)	-0.0352	(72.55)	-0.1569	(13.75)
AR-CAPM	**0.2546**	**(0.09)**	**0.2182**	**(0.59)**	0.1512	(7.34)	-0.0371	(63.46)	-0.0744	(34.47)	-0.1389	(10.86)
MPPM(3)	0.1232	(24.06)	**0.2290**	**(3.88)**	0.0640	(52.66)	-0.1713	(11.19)	-0.0626	(57.12)	**-0.2288**	**(3.42)**

This Table reports the performance of the differences portfolios between the top performing funds (D10) and the bottom performing funds (D1) considering the different risk-adjusted performance measures. Portfolios are rebalanced monthly. Results with transaction costs of 0.25% and 2% are reported. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model. *Ex post* performance is measured by excess returns, Fama and French (2015) five-factor alphas and Carhart (1997) four-factor alphas. Excess returns, alphas and their corresponding p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.

Table 11 – Differences in *ex post* performance between the top (T3) and the bottom (T1) performing funds: monthly rebalancing

	Equally weighted portfolios						Value weighted portfolios					
	T3-T1 Mean excess returns		T3-T1 Fama and French (2015) five-factor alphas		T3-T1 Carhart (1997) four-factor alphas		T3-T1 Mean excess returns		T3-T1 Fama and French (2015) five-factor alphas		T3-T1 Carhart (1997) four-factor alphas	
Sharpe	**0.2043**	**(0.07)**	**0.1634**	**(1.20)**	0.1316	(5.54)	**0.1539**	**(1.14)**	**0.1273**	**(4.56)**	0.0995	(15.12)
ERVaR-HS(0.05)	**0.1980**	**(0.12)**	**0.1526**	**(2.30)**	0.1241	(7.38)	**0.1746**	**(0.69)**	0.1291	(6.28)	0.1132	(10.31)
ERVaR-NG(0.05)	**0.2101**	**(0.08)**	0.1224	(7.39)	**0.1474**	**(1.65)**	**0.1846**	**(0.30)**	0.0983	(15.36)	**0.1285**	**(3.10)**
ERCVaR-HS(0.05)	**0.1870**	**(0.27)**	**0.1628**	**(1.67)**	0.1093	(13.08)	**0.1586**	**(1.31)**	**0.1388**	**(4.22)**	0.0957	(18.41)
ModSortino	**0.1922**	**(0.16)**	**0.1709**	**(0.89)**	0.1170	(9.68)	**0.1490**	**(1.48)**	**0.1370**	**(2.93)**	0.0877	(20.69)
FT(1,1)	**0.2004**	**(0.09)**	**0.1666**	**(1.13)**	0.1233	(8.24)	**0.1598**	**(1.03)**	**0.1335**	**(3.43)**	0.0995	(15.72)
FT(1,2)	**0.1793**	**(0.64)**	**0.1796**	**(0.94)**	0.1026	(16.07)	**0.1487**	**(2.34)**	**0.1456**	**(2.73)**	0.0837	(23.88)
RR(0.05,0.05)	0.0155	(74.96)	0.0868	(10.06)	-0.0155	(80.00)	-0.0200	(70.46)	0.0552	(33.21)	-0.0506	(43.34)
Calmar	**0.1659**	**(1.07)**	**0.1850**	**(0.41)**	0.1077	(12.81)	0.1100	(8.85)	**0.1337**	**(3.30)**	0.0657	(31.68)
Sterling	**0.1961**	**(0.14)**	**0.1598**	**(1.72)**	0.1236	(7.79)	**0.1627**	**(1.06)**	**0.1299**	**(4.92)**	0.1006	(14.58)
Burke	**0.1869**	**(0.24)**	**0.1650**	**(1.31)**	0.1136	(11.08)	**0.1391**	**(2.83)**	0.1187	(7.74)	0.0787	(26.66)
AR-CAPM	**0.2186**	**(0.01)**	**0.1747**	**(0.75)**	0.1212	(6.55)	**0.1543**	**(0.31)**	0.1163	(5.41)	0.0789	(20.44)
MPPM(3)	0.1294	(5.77)	**0.2025**	**(0.58)**	0.0829	(21.98)	0.1104	(12.29)	**0.1816**	**(1.45)**	0.0613	(36.68)

This Table reports the performance of the differences portfolios between the top performing funds (T3, representing the top 30% performing funds) and the bottom performing funds (T1, representing the bottom 30% performing funds) considering the different risk-adjusted performance measures. Portfolios are rebalanced monthly. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model. *Ex post* performance is measured by excess returns, Fama and French (2015) five-factor a and Carhart (1997) four-factor alphas. Excess returns, alphas and their corresponding p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.

Table 12 – Differences in *ex post* performance between the top (D10) and the bottom (D1) performing funds: annually rebalancing

	Equally weighted portfolios						Value weighted portfolios					
	D10-D1 Mean excess returns		D10-D1 Fama and French (2015) five-factor alphas		D10-D1 Carhart (1997) four-factor alphas		D10-D1 Mean excess returns		D10-D1 Fama and French (2015) five-factor alphas		D10-D1 Carhart (1997) four-factor alphas	
Sharpe	0.1449	(11.16)	0.1057	(26.74)	**0.1589**	**(4.86)**	0.0524	(59.53)	0.0326	(76.82)	0.1154	(13.29)
ERVaR-HS(0.05)	0.1264	(17.17)	0.0544	(60.15)	0.1273	(13.85)	0.0237	(82.86)	-0.0533	(69.22)	0.0834	(33.28)
ERVaR-NG(0.05)	0.1600	(5.44)	0.0578	(52.68)	**0.1600**	**(4.72)**	0.0993	(23.59)	0.0155	(85.75)	**0.1313**	**(4.87)**
ERCVaR-HS(0.05)	0.1446	(13.29)	0.1304	(20.20)	**0.1735**	**(4.66)**	0.0781	(41.82)	0.0978	(32.03)	0.1336	(8.27)
ModSortino	0.1289	(17.96)	0.1127	(26.24)	0.1587	(6.27)	0.0425	(68.84)	0.0481	(66.25)	0.1125	(16.76)
FT(1,1)	0.1215	(19.43)	0.0720	(47.40)	0.1372	(8.80)	0.0138	(89.58)	-0.0150	(90.18)	0.0794	(32.06)
FT(1,2)	0.1615	(10.11)	0.1644	(9.96)	**0.1949**	**(3.67)**	0.1072	(31.99)	0.1255	(20.43)	0.1560	(7.20)
RR(0.05,0.05)	-0.0561	(47.42)	0.0052	(94.20)	0.0145	(84.70)	-0.0108	(90.01)	0.0517	(45.14)	0.0282	(72.35)
Calmar	0.1528	(11.75)	0.1533	(12.43)	**0.1816**	**(3.64)**	0.0722	(51.59)	0.1305	(18.88)	0.1408	(11.02)
Sterling	0.1412	(13.36)	0.0977	(33.96)	0.1632	(5.66)	0.0447	(65.79)	0.0233	(84.20)	0.1145	(15.97)
Burke	0.1288	(18.53)	0.1045	(31.54)	0.1596	(6.74)	0.0352	(74.11)	0.0254	(82.50)	0.1066	(19.20)
AR-CAPM	0.1464	(5.38)	0.0803	(35.73)	**0.1455**	**(1.88)**	0.0876	(19.68)	0.0432	(54.55)	0.1040	(5.07)
MPPM(3)	0.0374	(70.52)	0.1281	(13.23)	0.1295	(15.25)	0.0217	(84.41)	0.1244	(19.12)	0.1412	(12.98)

This Table reports the performance of the differences portfolios between the highest performing funds (D10) and the bottom performing funds (D1) considering the different risk-adjusted performance measures. Portfolios are rebalanced annually. ERVaR-HS and ERCVaR-HS are computed on the basis of a VaR and CRVaR using historical returns and numerical integration; and ERVaR-NG is computed based on parametric VaR using a normal-GARCH model.*Ex post* performance is measured by excess returns, Fama and French (2015) five-factor alphas and Carhart (1997) four-factor alphas. Excess returns, alphas and their corresponding p-values are expressed in percentage. Values reported in bold are statistically significant at the 5% level.