C.I.R.P.É.E

Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **13-29**

The Peer Performance of Hedge Funds

David Ardia
Kris Boudt

Septembre/September 2013

Ardia : Corresponding author. Département de finance, assurance et immobilier, Université Laval, Québec City (Québec), Canada; and CIRPÉE
david.ardia@fsa.ulaval.ca
Boudt : Solvay Business School, Vrije Universiteit Brussel; and Faculty of Economics and Business, VU University Amsterdam, The Netherlands
kris.boudt@vub.ac.be

Work in progress; comments and welcome. We are grateful to Peter Carl, Michel Dubois, Ivan Guidotti, Lennart Hoogerheide, Simon Keel, Doug Martin, Enrico Schumann and participants at various conferences and seminars for useful comments. Financial support from aeris CAPITAL AG and the Dutch science foundation is gratefully acknowledged. The views expressed in this paper are the sole responsibility of the authors. Any remaining errors or shortcomings are the authors' responsibility.

Abstract:
An essential component in the analysis of (hedge) fund returns is to measure its performance with respect to the group of peer funds. Through the analysis of risk-adjusted return percentiles an answer is given to the question how many funds are out-performed by the focal fund. In case all funds perform equally well, this approach will lead a random number between zero and one, depending on how lucky the fund is. We use the false discovery rate approach to construct relative performance ratios that account for the uncertainty in estimating the performance differential of two funds. Our application is on hedge funds, which leads us to develop a test for equality of the modified Sharpe ratio of two funds. The effectiveness of the method is illustrated with a Monte Carlo study and an empirical study is performed on the Hedge Fund Research database.

Keywords: Equal-performance ratio, false discovery rate, hedge fund, modified Sharpe ratio, out-performance ratio, peer group, performance measurement

JEL Classification: C12, C21, C22

1. Introduction

> *Morningstar uses the Morningstar Category as the primary peer group for a number of calculations, including percentile ranks, fund-versus-category-average comparisons, and the Morningstar Rating™.*
>
> (The Morningstar Rating™ methodology)

How does a fund rank compared to its peers in terms of performance? The industry standard to answer this question is to study the fund's rank and estimate the percentage of funds that are out-performed by the focal fund as the percentage of funds for which the performance measure of the focal fund is higher. We call this a *naive* approach to peer fund analysis, as it completely ignores the estimation uncertainty in performance measures, such as the (modified) Sharpe ratio or Jensen's alpha.

A less naive solution would be to calculate the percentage of funds for which the performance of the focal fund is *significantly* higher, say at a 10% level. However, also this measure is not acceptable as it can be biased for two reasons. First, because of the multiple testing on a large number of peer funds, we have that under the null of no significant difference in talent, the estimated percentage of out-performance is on average 10%, while it truly is 0%. Second, in the case where the focal fund is truly out-performing its peers, it might be that because of bad luck, the test statistic lies in the region of non-rejection. The same problems arise when testing for the significance of Jensen's alpha of single funds.

The "Out-performance Ratio" that we propose uses the false discovery rate (FDR) approach by Storey (2002) to avoid these problems. It is inspired by the earlier work of Barras et al. (2010) in which the FDR methodology is used to estimate the percentage of truly skilled funds, which requires performance evaluation of individual funds. The building block of the proposed out-performance ratio is the p-value of a two-sided test of equal performance between the focal fund, on the one hand side, and the funds in the peer group, on the other hand.

This two-sided test can take many forms, as long as it is sufficiently powerful such that p-values on pairs of funds with truly different performance tend to be small. Ledoit and Wolf (2008) recommend a bootstrap method to test for the equal Sharpe ratio performance of two funds, accounting for the finite sample properties of the return distribution and the potential autocorrelation and heteroskedasticity. If one of the funds has non-normally distributed returns, comparing funds based on the Sharpe ratio is often not enough, as it ignores investors' positive preferences for odd moments (mean, skewness) and aversion to even moments (variance, kurtosis). This weakness of the Sharpe ratio is well known and several alternatives have been proposed. For the analysis of hedge fund returns, the modified Sharpe ratio proposed by Favre and Galeano (2002) and Gregoriou and Gueyie (2003) is now increasingly popular. It is defined as the ratio between the excess return of the hedge fund and its modified Value-at-Risk. The latter is an estimator for Value-at-Risk based on the Cornish-Fisher expansion and the first four moments of the return distribution. A statistical test for equal modified Sharpe ratio performance is missing. We fill this gap, and, illustrate its usefulness to analyze performance of hedge funds through an extensive Monte Carlo simulation and empirical application on the Hedge Fund Research (HFR) database.

The remainder of the article is organized as follows. Section 2 introduces the equal-performance ratio's methodology. Section 3 derives the asymptotic and bootstrap tests for the equality of modified Sharpe ratio of two funds. In Sections 4 and 5 we report the results of Monte Carlo study investigating the finite-sample performance of the proposed estimators. Section 6 illustrates the application of the proposed peer performance measures to analyze the relative performance of Equity Hedge, Event-Driven, Relative Value and Macro hedge funds. An accompanying R package is available from the authors' website.[1] Section 7 concludes and sketches directions for further research.

2. The peer performance ratios

We consider a universe with a total of $n + 1$ funds, where for each we can compute a risk-adjusted performance measure. This measure can take various forms, as long as there exists a valid two-sided test for equal performance of two funds. In practice, it can be e.g. the alpha of a factor model or, as we will see later, the (modified) Sharpe ratio. We denote by $\Delta_{\bullet i-j}$ the (true) difference of this performance measure between fund i and j $(i \neq j)$ and Δ_{i-j} is the corresponding estimate. Throughout the paper, we use the subscript symbol $\bullet$ to distinguish population-values (such as parameters $\Delta_{\bullet i-j}$) from sample-based estimates (such as Δ_{i-j}).

We wish to estimate the percentage of funds that have equal, less or greater risk-adjusted performance than fund i. Denote by $\pi^0_{\bullet i}$ $(n^0_{\bullet i})$, $\pi^+_{\bullet i}$ $(n^+_{\bullet i})$ and $\pi^-_{\bullet i}$ $(n^-_{\bullet i})$ the proportion (number) of funds for which $\Delta_{\bullet i-j} = 0$, $\Delta_{\bullet i-j} > 0$ and $\Delta_{\bullet i-j} < 0$, respectively.

2.1. The equal performance ratio

The building block of the proposed equal–performance ratio estimating $\pi^0_{\bullet i}$ are the p-values p_{i-j} associated to a two-sided test of the null hypothesis $H_0 : \Delta_{\bullet i-j} = 0$, for $j = 1, \ldots, n$. A crucial feature of the proposed estimators will be the difference in distribution of the p-values p_{i-j} when $\Delta_{\bullet i-j} = 0$ versus $\Delta_{\bullet i-j} \neq 0$.

If the test is sufficiently powerful, there exists a threshold value λ such that almost surely the p-value of the two-sided equal performance test is less than λ if the two funds have a different performance:

$$(A1): \quad \mathbb{P}(p_{i-j} < \lambda \,|\, \Delta_{\bullet i-j} \neq 0) = 1\,. \tag{1}$$

The validity of this assumption will depend on the magnitude of $\Delta_{\bullet i-j}$, the test-statistic itself, the calculation of its p-value (asymptotic versus bootstrap methods) and the sample size.

The second assumption that we make is that, for the chosen threshold value $0 < \lambda < 1$, the probability of p_{i-j} exceeding λ, when $\Delta_{\bullet i-j} = 0$, is $1 - \lambda$:

$$(A2): \quad \mathbb{P}(p_{i-j} \geq \lambda \,|\, \Delta_{\bullet i-j} = 0) = 1 - \lambda\,. \tag{2}$$

[1] All computations are performed in the R statistical computing language (R Development Core Team, 2011) with the package **CompStrat** (Ardia and Boudt, 2013). We rely on compiled functions for the bootstrap methodology and parallelize the computations using the package **snow** (Tierney et al., 2012).

This assumption is satisfied when, for a given pair (i, j), the p-values are uniformly distributed under the null of equal performance. To see that this is a realistic assumption, suppose that the p-values can be computed as $p_{i-j} \equiv F_{\bullet i-j}(\Delta_{i-j})$, with $F_{\bullet i-j}(\cdot)$ the cumulative distribution of $\Delta_{\bullet i-j}$. Those p-values have by construction a uniform distribution, since:

$$\begin{aligned}\mathbb{P}(p_{\bullet i-j} < p_{i-j}) &= \mathbb{P}(p_{\bullet i-j} < p_{i-j}) \\ &= \mathbb{P}(F_{\bullet i-j}(\Delta_{\bullet i-j}) < p_{i-j}) \\ &= \mathbb{P}(\Delta_{\bullet i-j} > F^{-1}_{\bullet i-j}(1 - p_{i-j})) \\ &= 1 - F_{\bullet i-j}(F^{-1}_{\bullet i-j}(1 - p_{i-j})) = p_{i-j}\,.\end{aligned} \tag{3}$$

In practice, $F_{\bullet i-j}(\cdot)$ and hence the p-values are estimated and assumption A2 will only be approximately satisfied.

Under (A1) and (A2) we have that the expected number of p-values exceeding λ is $(1 - \lambda)n^0_{\bullet i}$:

$$\begin{aligned}\mathbb{E}\left(\sum_{j\neq i} I[p_{i-j} \geq \lambda]\right) =& \sum_{j\neq i} \mathbb{E}(I[p_{i-j} \geq \lambda] \,|\, \Delta_{\bullet i-j} = 0) \\ &+ \sum_{j\neq i} \mathbb{E}(I[p_{i-j} \geq \lambda] \,|\, \Delta_{\bullet i-j} \neq 0) \\ =& (1 - \lambda)\, n^0_{\bullet i}\,,\end{aligned} \tag{4}$$

where $I[\cdot]$ denotes the indicator function. Hence a natural estimator for $n^0_{\bullet i}$ is the number of estimated p-values exceeding λ, divided by $1 - \lambda$:

$$n^0_i \equiv \min\left\{\frac{\sum_{j\neq i} I[p_{i-j} \geq \lambda]}{1 - \lambda}, n\right\}\,. \tag{5}$$

The corresponding estimate for the proportion of equal performance is:

$$\pi^0_i \equiv \frac{n^0_i}{n}\,. \tag{6}$$

The estimation of parameters based on the extrapolation of the highest p-values of multiple tests goes back to the false discovery rate (FDR) approach proposed by Storey (2002), and to the estimation of the percentage of talented mutual funds in Barras et al. (2010).[2]

The accuracy of the equal performance ratio estimator depends on both the accuracy of the pair-

[2]Since the estimator π^0_i has the form of a sample average, we expect, that in most relevant cases, the estimator π^0_i is not only unbiased but also consistent for $\pi^0_{\bullet i}$. The proof of this requires a suitable law of large number allowing for the dependence in the p-values. Indeed, even though each p-value is uniformly distributed under the null hypothesis, because of the correlation across hedge funds and the comparison with a common peer, the p-values corresponding to the different tests $\Delta_{\bullet i-j} = 0$, for different i and j, are not uniform distributed. In case of too strong dependence, it may make the estimator inconsistent (e.g. when all p-values are identical). In our sample, the average correlation between pairs of p-values with a common fund is 0.045. Because the dependence is not so extreme, we conjecture the estimator will be consistent in the population corresponding to our sample.

wise test for differences in the risk-adjusted performance measures, and hence the number of time series observations T, as well as the the cross-sectional number of funds. To make the latter more clear. Suppose we have a population of n_i p-values obtained without estimation error, i.e. $p_{i-j} \equiv F_{\bullet i-j}(\Delta_{i-j})$ and define by n_i^λ the number of funds in this population with p-value larger than λ. Using a random sample of n_i draws from the population (with replacement), $S := \sum_{j\neq i} I[p_{i-j} \geq \lambda]$ follows a binomial distribution with variance $nn_i^\lambda(n_i^\lambda - n)$, and hence the asymptotic distribution of $\pi_i^0 = S/(n(1-\lambda))$ is:

$$\pi_i^0 \overset{d}{\sim} \mathcal{N}\left(\pi_{\bullet i}^0, \frac{n_i^\lambda(n-n_i^\lambda)}{n^3(1-\lambda)^2}\right), \tag{7}$$

for n large.

2.2. The out- and under-performance ratios

Given the estimate of the number of peer funds with equal performance, n_i^0, and the observed performance differences Δ_{i-j}, we then estimate the number of funds that are out-performed by the focal fund, i.e. $n_{\bullet i}^+$. The estimate is based on the fact that $n_{\bullet i}^+$ corresponds to the naive estimate of the number of peer funds that are significantly out-performed by the focal fund (at a one-sided confidence level β, i.e. the number of funds for which Δ_{i-j} exceeds the β-quantile of the distribution of Δ_{i-j} under the null hypothesis, and denoted q_{i-j}^β), adjusted for the number of funds for which $q_{i-j}^\beta > 0$ by chance (when $\Delta_{\bullet i-j} \leq 0$) and the funds for which $\Delta_{i-j} < q_{i-j}^\beta$ when $\Delta_{\bullet i-j} > 0$:

$$\begin{aligned}
n_{\bullet i}^+ &= \sum_{j\neq i} I[\Delta_{\bullet i-j} > 0] \\
&= \sum_{j\neq i} I[\Delta_{i-j} > q_{i-j}^\beta] - \sum_{j\neq i} I[\Delta_{i-j} > q_{i-j}^\beta \,|\, \Delta_{\bullet i-j} = 0] \\
&\quad - \sum_{j\neq i} I[\Delta_{i-j} > q_{i-j}^\beta \,|\, \Delta_{\bullet i-j} < 0] + \sum_{j\neq i} I[\Delta_{i-j} < q_{i-j}^\beta \,|\, \Delta_{\bullet i-j} > 0]\,.
\end{aligned} \tag{8}$$

Given n_i^0, we can infer that, the number of false positives if $\Delta_{\bullet i-j} > 0$ is $n_i^0(1-\beta)$. The choice of β is a trade-off. For increasing values of β, the number of false positives $\sum_{j\neq i} I[\Delta_{i-j} > q_{i-j}^\beta \,|\, \Delta_{\bullet i-j} < 0]$ will converge to zero, but the number of false negatives $\sum_{j\neq i} I[\Delta_{i-j} < q_{i-j}^\beta \,|\, \Delta_{\bullet i-j} > 0]$ will increase. Throughout the paper, we take $q_{i-j}^\beta = 0$, corresponding to $\beta = 0.5$.[3]

This leads to the following definition of the out-performance ratio of fund i:

$$\pi_i^+ \equiv \frac{1}{n} \max\left\{\sum_{j\neq i} I[\Delta_{i-j} \geq q_{i-j}^\beta] - n_i^0(1-\beta), 0\right\}. \tag{9}$$

[3]In small samples, the distribution of Δ_{i-j} may be asymmetric and $q_{i-j}^{0.5}$ could then be estimated by the bootstrap procedure described in Subsection 3.1.

Finally, the under-performance ratio of fund i is given by:

$$\pi_i^- \equiv 1 - \pi_i^0 - \pi_i^+, \tag{10}$$

which can be interpreted as the adjusted frequency of peer funds that out-perform the focal fund i. Indeed, note that in the most common case where $\sum_{j\neq i} I[\Delta_{i-j} \geq q_{i-j}^{\beta}] > n_i^0(1-\beta)$, then:

$$\pi_i^- = \frac{1}{n}\left(\sum_{j\neq i} I[\Delta_{i-j} \geq q_{i-j}^{\beta}] - n_i^0(1-\beta)\right). \tag{11}$$

2.3. Choice of performance measure

The first step in the implementation of the out-performance ratio is the choice of performance measure. Broadly speaking, the performance of a fund is today either measured unconditionally using only the return series of the fund or conditionally by comparing the fund returns with a set of risk factors.

Examples of the first approach are the Sharpe, modified Sharpe, Sortino and Treynor ratios of the fund (Sharpe, 1994; Gregoriou and Gueyie, 2003; Sortino and Price, 1994; Treynor and Black, 1973). The previously mentioned performance measures have in common that the differential in performance can be rewritten as the average difference of portfolio returns standardized by a risk measure:

$$\Delta_{i-j} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{r_{i,t}}{\text{risk}_i} - \frac{r_{j,t}}{\text{risk}_j}\right), \tag{12}$$

where $r_{i,t}$ and $r_{j,t}$ be the realized return of funds i and j at time t ($t = 1\ldots,T$). with risk_i the standard deviation, the modified value-at-risk, the target semideviation and the fund's beta for the Sharpe, modified Sharpe, Sortino and Treynor ratios, respectively.

Examples of the second category include the intercept of the least squares regression of the fund returns on the four Carhart factors or the Fung and Hsieh hedge fund risk factors (Carhart, 1997; Fung and Hsieh, 2004). Let $\mathbf{f}_t$ be the $K \times 1$ vector of risk factors at time t. Then Δ_{i-j} is the intercept of the OLS regression of $(r_{i,t} - r_{j,t})$ on $\mathbf{f}_t$:

$$(r_{i,t} - r_{j,t}) = \Delta_{i-j} + \boldsymbol{\beta}_{i-j}'\mathbf{f}_t + \varepsilon_{i-j,t}, \tag{13}$$

for $t = 1,\ldots,T$, where $\boldsymbol{\beta}_{i-j}$ denotes the $K \times 1$ vector of factor exposures and:

$$\Delta_{i-j} = \frac{1}{T}\left(\sum_{t=1}^{T}(r_{i,t} - r_{j,t}) - \boldsymbol{\beta}_{i-j}'\mathbf{f}_t\right). \tag{14}$$

For both categories, we thus have that Δ_{i-j} can be rewritten as a time series average of potentially serially dependent observations, such that suitable central limit theorems can be invoked to obtain the (asymptotic) normal distribution of Δ_{i-j} and calculate its p-values. In practice, the common history of two fund returns is often short and bootstrap methods are expected to yield improved inference. In the sequel, we focus on one performance measure, namely the modified Sharpe ratio, but the calculation of the proposed peer performance ratio's is generally applicable to most of the commonly used performance measures.

2.4. Choice of λ

Based on Storey (2002), Barras et al. (2010, footnote 10) propose a bootstrap procedure to determine in a pure data driven way the value of λ that minimizes the estimated mean squared error of π_i^0. Alternatively, they recommend as a rule of thumb to use $\lambda = 0.5$. In the simulation study and empirical application, we used the bootstrap procedure with $B = 500$ replications.

3. Testing the equality of modified Sharpe ratios

It is a stylized fact that the return distribution for many hedge funds is asymmetric and has heavy tails. A typical investor has positive preferences for odd moments (mean, skewness) and aversion to even moments (variance, kurtosis), as discussed by Scott and Horvath (1980). This explains the rising popularity of Value-at-Risk as a measure to evaluate the risk of hedge funds rather than the standard deviation. Value-at-risk is defined as the negative value of the hedge fund return such that lower returns will only occur with at most a preset probability level denoted α. Common choices for α are 1%, 2.5%, 5% and 10%. Because fund returns are often only available at the monthly frequency and the higher estimation uncertainty of modfied VaR for smaller values of α (Boudt et al., 2008), especially in small samples, we will use $\alpha = 10\%$ in the application.

Of course, when estimating the Value-at-Risk, the non-normality of the return distribution should be accounted for. In many cases, the return distribution is unknown. A popular semi-parametric approach is then to approximate the true (unknown) distribution with the second order Cornish-Fisher expansion, i.e. the normal distribution plus some correction terms that account for the skewness and excess kurtosis of the return distribution (Cornish and Fisher, 1937). The resulting risk measure is called modified VaR, and was proposed by Zangari (1996). The modified VaR estimator is usually constructed by replacing the population moments with the corresponding sample moments. The modified VaR owes its popularity in practical work to its precision and, especially, to its explicit form, which makes it straightforward to compute and interpret. In our application, it allows for an explicit derivation of the standard error of the test statistic of equal modified Sharpe ratio performance.

The definition of modified VaR requires the first four moments. Let R_i be the return of hedge fund manager i. Let $m_{\bullet i}$ be the population mean for hedge fund manager i. Denote the q-th centered portfolio moment $m_{\bullet q,i} \equiv \mathrm{E}\left[(R_i - m_{\bullet,i})^q\right]$. For a loss probability α, the modified VaR estimate is:

$$\mathrm{mVaR}_{\bullet i}(\alpha) \equiv -m_{\bullet i} + \sqrt{m_{\bullet 2,i}}\left[-z_\alpha - \frac{1}{6}(z_\alpha^2 - 1)s_{\bullet i} - \frac{1}{24}(z_\alpha^3 - 3z_\alpha)k_{\bullet i} + \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)s_{\bullet i}^2\right], \tag{15}$$

with $s_{\bullet i}$ and $k_{\bullet i}$ the skewness and excess kurtosis:

$$s_{\bullet i} \equiv \frac{m_{\bullet 3,i}}{m_{\bullet 2,i}^{3/2}}, \qquad k_{\bullet i} \equiv \frac{m_{\bullet 4,i}}{m_{\bullet 2,i}^{2}} - 3, \tag{16}$$

and where z_α denotes the α-percentile of the standard normal distribution. Based on the large acceptance of modified VaR as a relevant hedge fund risk management tool, Gregoriou and Gueyie (2003) proposed the modified Sharpe ratio as a more suitable tool for assessing the hedge fund performance than the usual Sharpe ratio. The modified Sharpe ratio is defined as the ratio between the excess return of the hedge fund and its modified Value-at-Risk:

$$\mathrm{mSR}_{\bullet i}(\alpha) \equiv \frac{m_{\bullet i} - r_f}{\text{mVaR}_{\bullet i}(\alpha)}, \tag{17}$$

where r_f is the average risk-free rate at the corresponding horizon.

3.1. Pairwise test for equality of modified Sharpe ratios

Let $r_{i,t}$ and $r_{j,t}$ be the realized return of funds i and j at time t ($t = 1\ldots,T$). Replacing the population moments with sample moments in (15) and (17) provides us with estimates of the modified VaR and the modified Sharpe ratio. Denote these sample moments as m_i, $m_{2,i}$,$m_{3,i}$, $m_{4,i}$, s_i and k_i, and the resulting modified VaR and Sharpe ratio as $\text{mVaR}_i(\alpha)$ and $\mathrm{mSR}_i(\alpha)$.

We are interested in testing the null hypothesis of equal modified Sharpe ratios for fund i and j:

$$H_0 : \Delta_{\bullet i-j} \equiv \mathrm{mSR}_{\bullet i}(\alpha) - \mathrm{mSR}_{\bullet j}(\alpha) = \frac{m_{\bullet i}}{\text{mVaR}_{\bullet i}(\alpha)} - \frac{m_{\bullet j}}{\text{mVaR}_{\bullet j}(\alpha)} = 0, \tag{18}$$

based on the observed difference in the cross-product of the fund's return and the other fund's mVaR:

$$\Delta_{i-j} \equiv m_i \, \text{mVaR}_j(\alpha) - m_j \, \text{mVaR}_i(\alpha). \tag{19}$$

The test statistic we recommend to use for evaluating the null hypothesis is the ratio between Δ_{i-j} and its standard error. Under regularity conditions, the Delta method implies that this studentized test statistic is asymptotically normally distributed. However, because of the rather small samples for which hedge fund return data are typically available, we follow the proposal of Ledoit and Wolf (2008) and use a bootstrap methods for calculating the appropriate p-values, accounting for the finite sample size and temporal dependence in the data.

Standard error of Δ_{i-j}. To compute the standard error of Δ_{i-j}, we first express all centered moments as uncentered moments $g_{q,i} \equiv \frac{1}{T}\sum_{t=1}^{T} r_{i,t}^{q}$:

$$m_{2,i} = g_{2,i} - m_i^2 \tag{20}$$

$$m_{3,i} = g_{3,i} - 3m_i \, g_{2,i} + 2m_i^3 \tag{21}$$

$$m_{4,i} = g_{4,i} - 4m_i \, g_{3,i} + 6m_i^2 g_{2,i} - 3m_i^4. \tag{22}$$

It follows that Δ_{i-j} is a function of the mean and second to fourth moment about the origin of the two series:

$$\Delta_{i-j} \equiv f(m_i, g_{2,i}, g_{3,i}, g_{4,i}, m_j, g_{2,j}, g_{3,j}, g_{4,j})\,. \tag{23}$$

If Ψ_{i-j} is a consistent estimator of the asymptotic covariance matrix between these arguments, then an estimate for the standard error of Δ_{i-j} is:

$$s(\Delta_{i-j}) \equiv \sqrt{\frac{\nabla'_{i-j}\Psi_{i-j}\nabla_{i-j}}{T}}\,, \tag{24}$$

with ∇_{i-j} the gradient of f with respect to $(m_i, g_{2,i}, g_{3,i}, g_{4,i}, m_j, g_{2,j}, g_{3,j}, g_{4,j})'$. In this case, ∇_{i-j} is given by:

$$\nabla_{i-j} \equiv (\text{mVaR}_j\, dm_i - m_i d\,\text{mVaR}_j) - (\text{mVaR}_i\, dm_j - m_j d\,\text{mVaR}_i)\,, \tag{25}$$

with:

$$\begin{aligned} d\,\text{mVaR}_i \;=& - dm_i - z_\alpha \frac{1}{2\sqrt{m_{2,i}}} dm_{2,i} \\ & - \frac{1}{6}(z_\alpha^2 - 1)\left(\frac{1}{2\sqrt{m_{2,i}}} s_i dm_{2,i} + \sqrt{m_{2,i}} ds_i\right) \\ & - \frac{1}{24}(z_\alpha^3 - 3z_\alpha)\left(\frac{1}{2\sqrt{m_{2,i}}} k_i dm_{2,i} + \sqrt{m_{2,i}} dk_i\right) \\ & + \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)\left(\frac{1}{2\sqrt{m_{2,i}}} s_i^2 dm_{2,i} + 2\sqrt{m_{2,i}} s_i ds_i\right)\,, \end{aligned} \tag{26}$$

and:

$$ds_i = \frac{m_{2,i}^{3/2} dm_{3,i} - 1.5 m_{3,i} m_{2,i}^{1/2} dm_{2,i}}{m_{2,i}^3} \tag{27}$$

$$dk_i = \frac{m_{2,i}^2 dm_{4,i} - 2m_{4,i} m_{2,i} dm_{2,i}}{m_{2,i}^4}\,, \tag{28}$$

and similarly for j. The gradient for the underlying moments is:

$$\begin{aligned}
dm_i &= (1,0,0,0,0,0,0,0)' && (29)\\
dm_j &= (0,0,0,0,1,0,0,0)' && (30)\\
dm_{2,i} &= (-2m_i,1,0,0,0,0,0,0)' && (31)\\
dm_{2,j} &= (0,0,0,0,-2m_j,1,0,0)' && (32)\\
dm_{3,i} &= (-3g_{2,i}+6m_i^2,-3m_i,1,0,0,0,0,0)' && (33)\\
dm_{3,j} &= (0,0,0,0,-3g_{2,j}+6m_j^2,-3m_j,1,0)' && (34)\\
dm_{4,i} &= (-4g_{3,i}+12m_i g_{2,i}-12m_i^3,6m_i^2,-4m_i,1,0,0,0,0)' && (35)\\
dm_{4,j} &= (0,0,0,0,-4g_{3,i}+12m_i g_{2,i}-12m_i^3,6m_i^2,-4m_i,1)'\,. && (36)
\end{aligned}$$

Alternative test statistic. An alternative way to test the difference in (23) is to consider directly the observed difference in modified Sharpe ratios:

$$\Delta_{i-j} \equiv \mathrm{mSR}_i(\alpha) - \mathrm{mSR}_j(\alpha)\,. \quad (37)$$

In this case, ∇_{i-j} is given by:

$$\nabla_{i-j} \equiv \left(\frac{\text{mVaR}_i\, dm_i - m_i d\,\text{mVaR}_i}{\text{mVaR}_i^2}\right) - \left(\frac{\text{mVaR}_j\, dm_j - m_j d\,\text{mVaR}_j}{\text{mVaR}_j^2}\right)\,, \quad (38)$$

with the expressions for m_i, mVaR_i, dm_i, and $d\,\text{mVaR}_i$ $(i = 1,\ldots,n)$ being the same.

The apparent advantage of this alternative test statistic is that its construction is similar as the traditional one used by Ledoit and Wolf (2008) to test for the equality of two Sharpe ratios. The disadvantage however is that both the numerator and denominator depend on the inverse of the modified Value-at-Risk. For funds with positive sample skewness, the modified Sharpe ratio can be close to zero, and extremely fat tailed. By considering the test statistic that cross-multiplies the mean and mVaR of the two funds, we obtain a test statistic that is analytically more simple and better behaved. We confirm this statement in the simulation study.

Estimation of the covariance matrix. For i.i.d. data we use the sample estimator of the covariance matrix. For time series data with possible dependence, we follow Ledoit and Wolf (2008) in using the heteroscedasticity and autocorrelation robust (HAC) kernel estimators of Andrews (1991) and Andrews and Monahan (1992).

Calculation of p-values. Because of the small sample size in our empirical application, we use the bootstrap method to compute the p-values of the test statistics under the null of equal Sharpe ratios. To generate bootstrap data, we resample with replacement either individual pairs, as in the i.i.d. bootstrap of Efron (1979), or blocks of fixed size $l \geq 1$, following the circular block bootstrap of Politis and Romano (1992). Based on the B bootstrap pairs $(r_{t,i}^{*b}, r_{t,j}^{*b})$ the bootstrap test statistics are computed:

$$t^{*b}_{i-j} \equiv \frac{\Delta^{*b}_{i-j}}{s(\Delta^{*b}_{i-j})}\,, \tag{39}$$

where $*b$ denotes the estimators computed on the b-th bootstrap data set. Since the distribution of the test statistic may be asymmetric in small samples, we follow Barras et al. (2010) in computing the p-value as:

$$p_{i-j} \equiv 2 \times \min\left(\frac{1}{B}\sum_{b=1}^{B} I[t^{*b}_{i-j} > t_{i-j}], \frac{1}{B}\sum_{b=1}^{B} I[t^{*b}_{i-j} < t_{i-j}]\right)\,, \tag{40}$$

where $I[t^{*b}_{i-j} > t_{i-j}]$ is an indicator function that takes the value of one if the bootstrap test statistic t^{*b}_{i-j} is higher than the estimated test-statistic t_{i-j}.

Choice of peer funds. Hoberg et al. (2013) review the two main approaches towards specifying peer funds. One approach uses historical returns and obtain the peer funds through cluster analysis of the returns or an analysis of the coefficients in a regression of fund returns on benchmark indices. The second approach classifies funds based on the fund holdings.

4. Size and power of pairwise test for equal modified Sharpe ratios

Ledoit and Wolf (2008) perform an extensive study on the size properties of different tests for equal Sharpe ratios and recommend the use of a studentized time series bootstrap approach. In Subsection 4.1 we do a similar size study for the proposed test on equal modified Sharpe ratio performance (at the 90% risk level, i.e. $\alpha = 10\%$). In particular, we provide evidence in favor of the particular test statistic using the cross-product of average return and modified Value-at-Risk in (19) rather than the difference in the sample-based modified Sharpe ratios in (38). Then, in Subsection 4.2 we document the similarity in power of the proposed modified Sharpe ratio test and the Sharpe ratio test of Ledoit and Wolf (2008) in case of normality and illustrate the impact of skewness and kurtosis. In the simulation studies, we set the number of Monte Carlo replications to five hundred. For the bootstrap test we set $B = 500$.

4.1. Modified Sharpe - Size

In the scenarios testing the size of the different tests, we need that the null hypothesis of equal (modified) Sharpe ratios is true and achieve this by generating the returns of fund i and j from an identical marginal return process, that are joined together in a bivariate distribution using a Gaussian copula with correlation of 0.5.

The marginal return process have an unconditional mean and variance equal to one. They are either (i) i.i.d., or (ii) follow an AR(1) model with auto-regressive parameter equal to 0.4 or (iii) a GARCH model with parameters $\omega = 0.20$, $\alpha = 0.10$ and $\beta = 0.70$. The conditional marginal innovations are either: (i) normal, (ii) standardized Student-t with six degrees of freedom or (iii) standardized skewed Student-t with six degrees of freedom and asymmetry parameter $\xi = 0.75$ (negative skewness). The standardized skewed Student-t distribution is the one proposed by Fernández and Steel (1998) and Lambert and Laurent (2001). Its density function has two parameters: ξ and ν. The parameter $\xi > 0$ is defined as the square root of the ratio of probability masses above and below the mode of the distribution such that the sign of $\log \xi$ indicates the sign of skewness. The parameter $\nu > 0$ models the tail thickness. Holding ξ fixed, we have that the smaller ν is, the thicker the tails are. When $\xi = 1$, the standardized skewed Student-t distribution coincides with the standardized Student-t distribution, and for $\nu \to \infty$, the standard Gaussian distribution is the limiting case. Overall, this leads to nine data generating processes. The size of the simulated DPGs is set to $T = 60$ (five years of monthly data as this is typical for hedge funds data) and $T = 500$ to see the asymptotic properties of the estimators.

For each DGP, we report in Figure 1 the 95% confidence bands of the rejection rate using four different test statistics at a significance level of 5%:

- [blue solid]: Mean-modified VaR (at the 90% risk level)cross-product test statistic in (19), with critical values from the asymptotic normal with HAC standard errors;
- [blue dashed]: Difference in mean-modified VaR (at the 90% risk level) ratios test statistic in (38), with critical values from the asymptotic normal with HAC standard errors;
- [red solid]: Block-bootstrap implementation of mean-modified VaR (at the 90% risk level) cross-product test statistic in (19);
- [red dashed]: Block-bootstrap implementation of mean-modified VaR (at the 90% risk level) ratios test statistic in (38).

In the block-bootstrap, the block size have been defined with the automatic block-length approach by Politis and White (2004). As the methodology is computationally costly, we have taken the average block-length obtained over ten Monte Carlo replications. For the i.i.d. DGP we set the block-size to one. For the AR model, to block-size 15 (for $T = 60$) and 20 (for $T = 500$), and for the GARCH to block-size 20 (for $T = 60$) and 30 (for $T = 500$). Alternative approaches such as the $T^{1/3}$ rule of thumb yield slightly more liberal results.

In order to save space, we report size results at the 5% level and using the 90% risk level in the modified VaR calculation; results at the 10% levels and/or 95% risk level are qualitatively similar and are available from the authors upon request. We verified also the size properties of the Sharpe ratio test and found confirmation of the good size properties of the Sharpe ratio test using HAC standard errors and bootstrapped p-values, as in Ledoit and Wolf (2008). The Sharpe ratio test has correct size both in the case of using the difference in Sharpe ratios and the difference in the cross-product of returns and volatility as a test statistic. We will therefore use in the application and power analyis of the next section, the standard implementation of the equal Sharpe ratio test using the ratio test statistic.

For the equal modified Sharpe ratio test, the conclusions are different. The first important result that we can see in Figure 1 is that the test statistic using the difference in mean-modified VaR ratios is not reliable. If we calculate its p-values under the normal distribution with HAC standard errors, then the test statistic is too conservative. For all nine DGP's considered, the 95% confidence interval for the percentage number of rejections is systematically below 5%. When the p-values are computed using the block bootstrap approach, then the test is too liberal and rejection frequencies tend to be above 20%.

A second finding is that the test statistic based on the cross-product of the average return and modified VaR of the two funds is slightly oversized in small samples and, except for the GARCH cases, it has correct size in the large sample case. For the DGPs with GARCH dynamics in the volatility, the test is oversized. One should not worry too much about this, as in applications on monthly (hedge) fund returns, the GARCH dynamics (if any) are less pronounced than the ones simulated here (see e.g. Barras et al., 2010, footnote 15).

Overall, we can thus conclude that the test statistic based on the cross-product of the average return and modified VaR of the two funds is certainly preferable to the one using the difference of ratios. In small samples, it tends to be slightly too liberal.

In the next subsection, we analyze the power of the modified Sharpe ratio test using the recommended implementation with the cross-product test statistic and compare it with the power of the usual Sharpe ratio test.[4]

[4]We did the same analysis for the Sharpe test and found similar size properties for the test statistic based on the ratio and the cross-product.

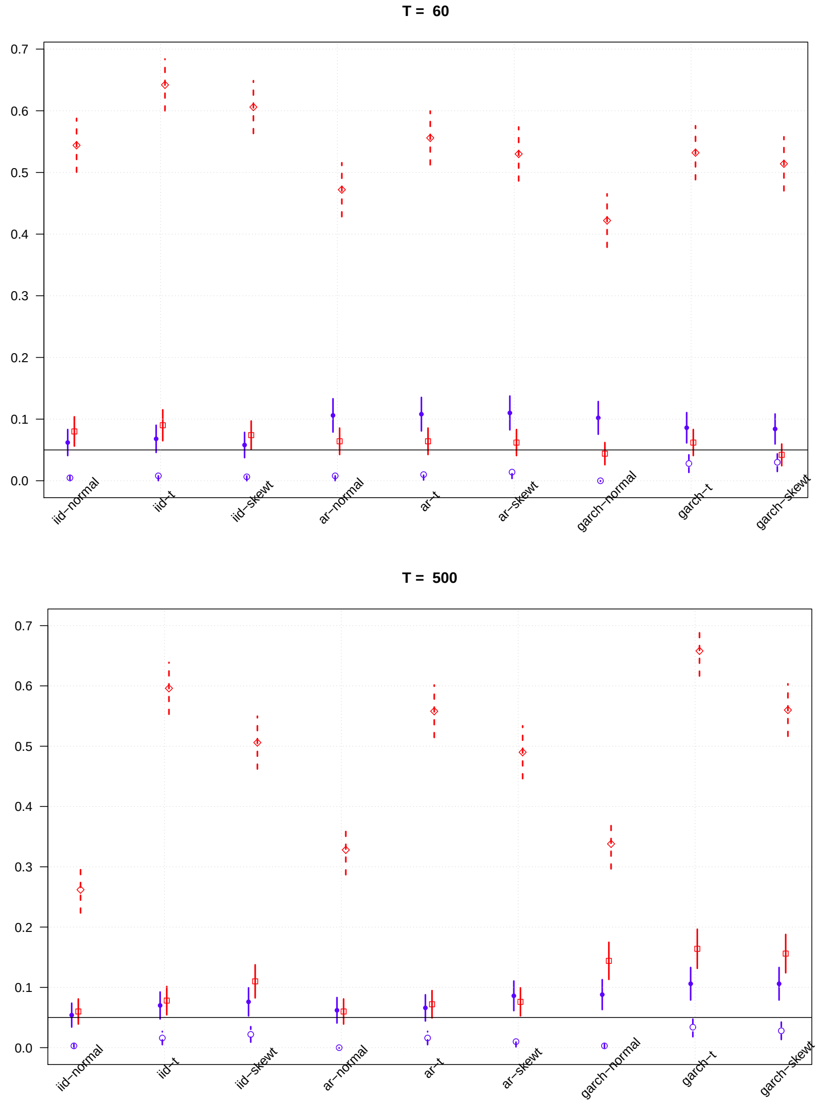


Figure 1: Monte Carlo analysis: size results for the modified Sharpe ratio (at the 90% risk level, i.e. $\alpha = 10\%$). The bars represent the 95% confidence bands of the frequency of rejecting the null hypothesis at the 5% significance level (over the five hundred Monte Carlo replications). The first two bands are for the HAC test (in blue, product in solid line and quotient in dashed line) while the last two bands are for the block-bootstrap (in red, product in solid line and quotient in dashed line). The DGPs are either (i) i.i.d., (ii) AR(1) or (iii) GARCH with innovations being either (i) normal, (ii) Student-t or (iii) skewed Student-t. Marginals are linked by a Gaussian copula with correlation coefficient set at 0.5.

4.2. *Modified Sharpe - Power*

The previous subsection confirmed the good size properties of the test for equality of modified Sharpe ratios of the focal fund i and the alternative peer fund j. In Figures 2 and 3 we study the power of the test when the difference in Sharpe ratio between the two funds increases. As a reference point, we take for fund i the returns from a skewed Student-t distribution with mean $\mu = 0.445\%$, standard deviation $\sigma = 1.07\%$, skewness parameter $\xi = 1$ (no skewness) and degrees of freedom parameter $\nu = 500$ (close to normal tails). This parameter setup corresponds to one of the out-performing funds in our empirical database. We consider the following variations to test the power of the test statistics:

SC1 $\mu_i = 0.445\%$, $\sigma_i = \sigma_j = 1.07\%$, $\xi_i = \xi_j = 1$, $\nu_i = \nu_j = 500$ and, for several values of γ between 1 and 0.1, μ_j is such that the Sharpe ratio of fund j equals γ times the Sharpe ratio of fund i;

SC2 $\mu_i = 0.445\%$, $\sigma_i = \sigma_j = 1.07\%$, $\xi_i = 1$, $\xi_j = 0.75$, $\nu_i = \nu_j = 500$ and, for several values of γ between 1 and 0.1, μ_j is such that the Sharpe ratio of fund j equals γ times the Sharpe ratio of fund i;

SC3 $\mu_i = 0.445\%$, $\sigma_i = \sigma_j = 1.07\%$, $\xi_i = 1$, $\xi_j = 0.75$, $\nu_i = 500$, $\nu_j = 6$ and, for several values of γ between 1 and 0.1, μ_j is such that the Sharpe ratio of fund j equals γ times the Sharpe ratio of fund i;

SC4 $\mu_i = \mu_j = 0.445\%$, $\sigma_i = 1.07\%$, $\xi_i = \xi_j = 1$, $\nu_i = \nu_j = 500$ and, for several values of γ between 1 and 0.1, σ_j is such that the Sharpe ratio of fund j equals γ times the Sharpe ratio of fund i;

SC5 $\mu_i = \mu_j = 0.445\%$, $\sigma_i = 1.07\%$, $\xi_i = 1$, $\xi_j = 0.75$, $\nu_i = \nu_j = 500$ and, for several values of γ between 1 and 0.1, σ_j is such that the Sharpe ratio of fund j equals γ times the Sharpe ratio of fund i;

SC6 $\mu_i = \mu_j = 0.445\%$, $\sigma_i = 1.07\%$, $\xi_i = 1$, $\xi_j = 0.75$, $\nu_i = 500$, $\nu_j = 6$ and, for several values of γ between 1 and 0.1, σ_j is such that the Sharpe ratio of fund j equals γ times the Sharpe ratio of fund i;

Under SC1 en SC4 the (population) Sharpe ratio and modified Sharpe ratio of fund i and fund j are the same for $\gamma = 1$. For SC2, SC3, SC5 and SC6, the modified Sharpe ratios are different for funds i and j, and we expect to see a gain in power using the modified Sharpe ratio test rather than the Sharpe ratio test, already for $\gamma = 1$.

The power is computed as the average frequency of rejection of the null hypothesis at the 5% level. The solid blue lines with and without symbols in Figure 2 and 3 shows the evolution of the power of the modified Sharpe ratio and standard Sharpe ratio test when the two fund returns are normally distributed (SC1, SC4). We see that the power of the two tests is similar.

If we allow for positive skewness in fund i and keep close to normal tails (SC2, SC5), we see that, especially in large samples, the modified Sharpe ratio test is slightly more powerful.

In case of fat tails (SC3, SC6), there is a higher estimation uncertainty in the modified Sharpe ratio estimates, leading to a lower power of the modified Sharpe ratio test compared to the Sharpe ratio test.

Overall, we find that for increasing differences in the mean parameter μ and volatility parameter σ, the increase in power of both the equal Sharpe ratio test and modified Sharpe ratio test is rather slow. In the next subsection, we will see that the slope of the out-performance ratio is much more steep.

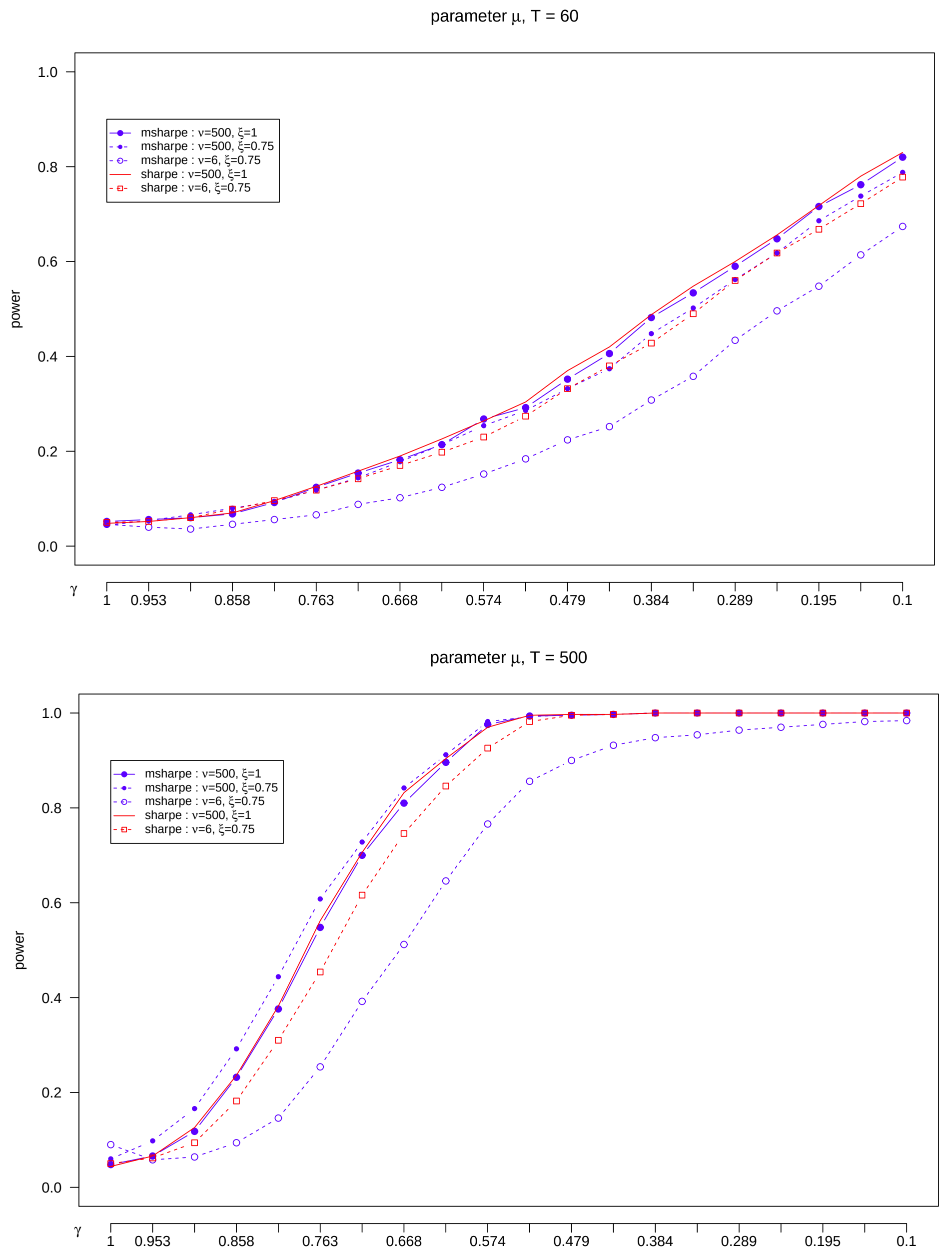
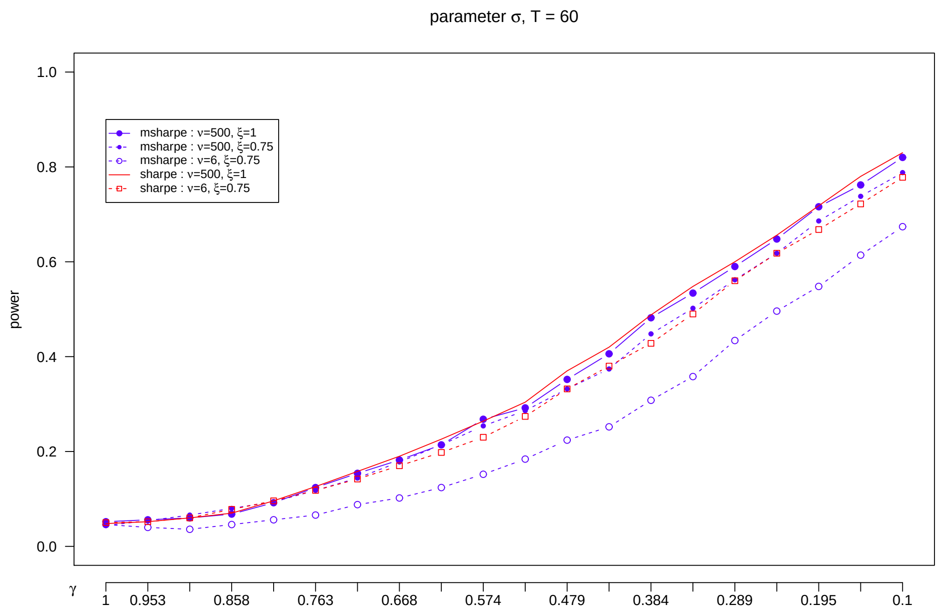


Figure 2: Monte Carlo analysis: power results when varying parameter μ for the modified Sharpe ratio (at the 90% risk level, i.e. $\alpha = 10\%$) and Sharpe ratio. The plot displays the frequency of rejections (over the five hundred Monte Carlo replications) at the 5% significance level of the null hypothesis of equal performance for deflated modified Sharpe funds, i.e. for γ ranging from 1 (equal performance) to 1/10 (under-performance). For each value of γ, μ is determined for the deflated fund. The draws from the focal fund and the alternative fund are generated using a skewed Student-t distribution with i) approximative normal ($\xi = 1, \nu = 500$) in solid blue line with bullet, ii) skewed normal ($\xi = 0.75, \nu = 500$) in dashed blue line with point, and iii) skewed fat-tailed ($\xi = 0.75, \nu = 6$) in dotted blue line with circle. Top: results for $T = 60$; Bottom: results for $T = 500$. The number of Monte Carlo replication is set to five hundred. For the modified Sharpe, the bootstrap test is used with block of size one and $B = 500$ bootstrap replications.



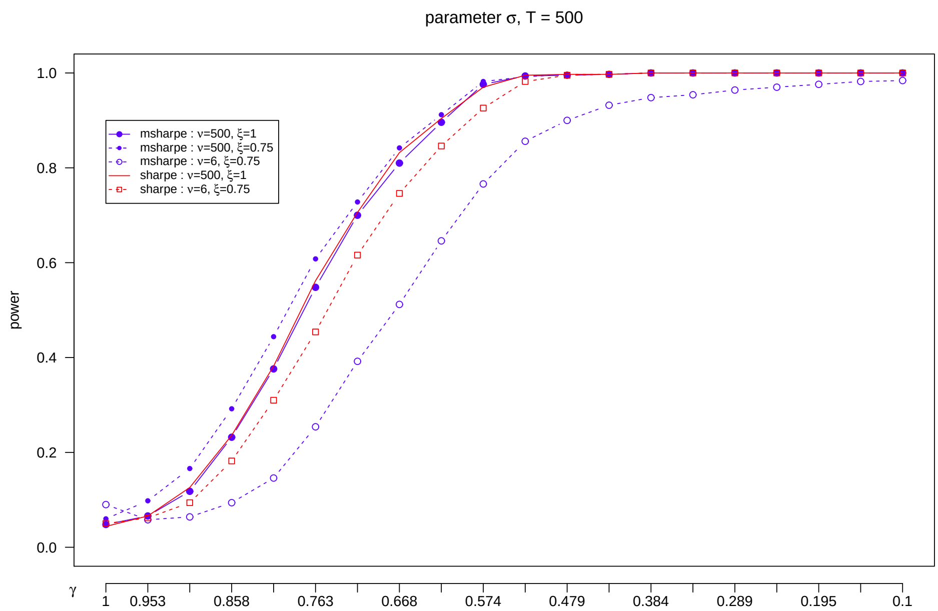


Figure 3: Monte Carlo analysis: power results when varying parameter σ for the modified Sharpe ratio (at the 90% risk level, i.e. $\alpha = 10\%$) and Sharpe ratio. For details on the Monte Carlo study see the caption of Figure 2.

5. Sensitivity of the out-performance ratio

The peer performance ratios for a focal fund i result from the aggregation of the p-values of the pairwise test of equal performance between fund i and j. Suppose fund i out-performs all peers. For the estimated out-performance ratio to be close to 100%, relatively large sample sizes and/or substantial large differences in true performance are needed.

We illustrate this in Figure 4 where we consider the first three scenarios of the power analysis in Subsection 4.2, with the only difference that the pair-wise test is performed between fund i and $n = 10$ peer funds j. We see that the out-performance ratio behaves as expected. It increases smoothly to one when we increase the differential in performance between fund i and its peer funds. Increasing the sample size increases sharply the speed of convergence. Importantly, from the comparison of the power curve in Figure 2 with the out-performance ratio curve in Figure 4, we see that the out-performance ratio curve is much more steep.

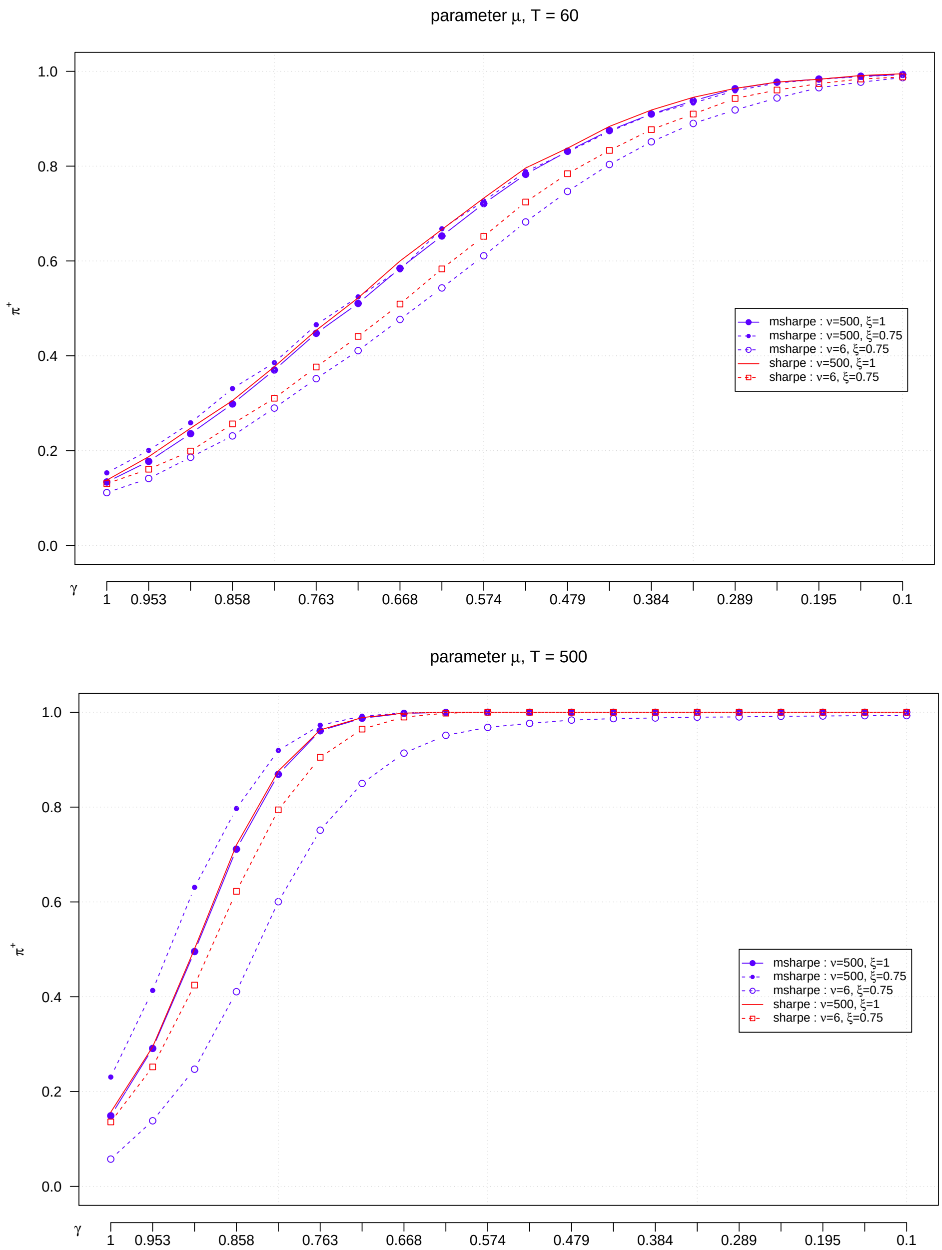


Figure 4: Monte Carlo analysis: out-performance ratio π^+ in (9) when varying parameter μ for the modified Sharpe ratio (at the 90% risk level, i.e. $\alpha = 10\%$) and Sharpe ratio. The number of peer funds is set to $n = 10$. The number of observations is set to $T = 60$ (top plot) and $T = 500$ (bottom plot). For details on the Monte Carlo study see the caption of Figure 2.

6. Empirical illustration

We now provide examples of diagnostic plots and tables that can be useful in the peer performance analysis of hedge funds in practice.

Our dataset is composed by two hundred funds taken from the HFR database for $T = 60$ monthly net returns computed over a period ranging from November 2006 to November 2011; see Guidotti and Nagy (2011) for the details.[5] We focus on the following strategies: Equity Hedge (50 funds), Event-Driven (50 funds), Relative Value (50 funds) and Macro (50 funds).

The main plot we recommend to use is in Figure 5. It displays in the left panel the monthly modified Sharpe ratio of the different funds (at the 90% risk level, i.e. $\alpha = 10\%$), sorted in descending order such that "Fund 1" has the highest modified Sharpe ratio. In the right panel, a barplot shows the estimated out-, equal and under-performance ratios in black, light gray and dark gray, respectively. These performance ratios are estimated taking all other hedge funds as the peer category. The equal performance ratio is obtained with the block-bootstrap test with $B = 500$ and block-size set to four (which is $\lceil T^{1/3} \rceil$); alternative specifications lead to similar results. The out- and under-performance ratios are obtained via the attributions (9) and (10).

It is also insightful to do the peer performance analysis relative to all funds belonging to the same hedge fund investment style. The resulting analysis could be presented in a similar plot as the one in Figure 5, but also in tabular form, as we show in Table 1. As in Figure 5, funds are sorted in descending order. The first column contains the name of the funds.[6] The second column shows the reported investment style. The third column and fourth columns report the percentile and the out- and under-performance compared to all funds. The fifth and sixth columns do the same, but with peer funds limited to be in the Equity Hedge investment style, and similarly for all other funds.

As an example, take fund AI. This fund pursues an Event-Driven investment style. This fund has the 9th highest modified Sharpe ratio of the 200 funds considered. Its estimated out-performance ratio is 76%, its under-performance ratio is 0% and hence its equal performance ratio is 24%. When its peer funds are limited to the Equity Hedge hedge funds, its estimated out-performance ratio is 100%. In contrast, when we the peer funds corresponds to the relative value style, its out-performance ratio is only 44% and its equal performance ratio is 66%.

Above, we have studied the relative performance of a fund compared to various universes defined by the hedge fund style. It is clear that some hedge fund styles tend to out-perform others. As an aggregate measure we propose to investigate in Table 2 the average relative performance ratios of a fund belonging to the category in row, compared to the funds belonging to the category in column. We notice that 36% of the Event-Driven strategies truly under-performed the Equity-Hedge strategies, while 55% of the Event-Driven strategies truly out-performed the Relative Value strategies.

[5] We are grateful to the authors for providing us with the dataset.

[6] Note that funds have been anonymized.

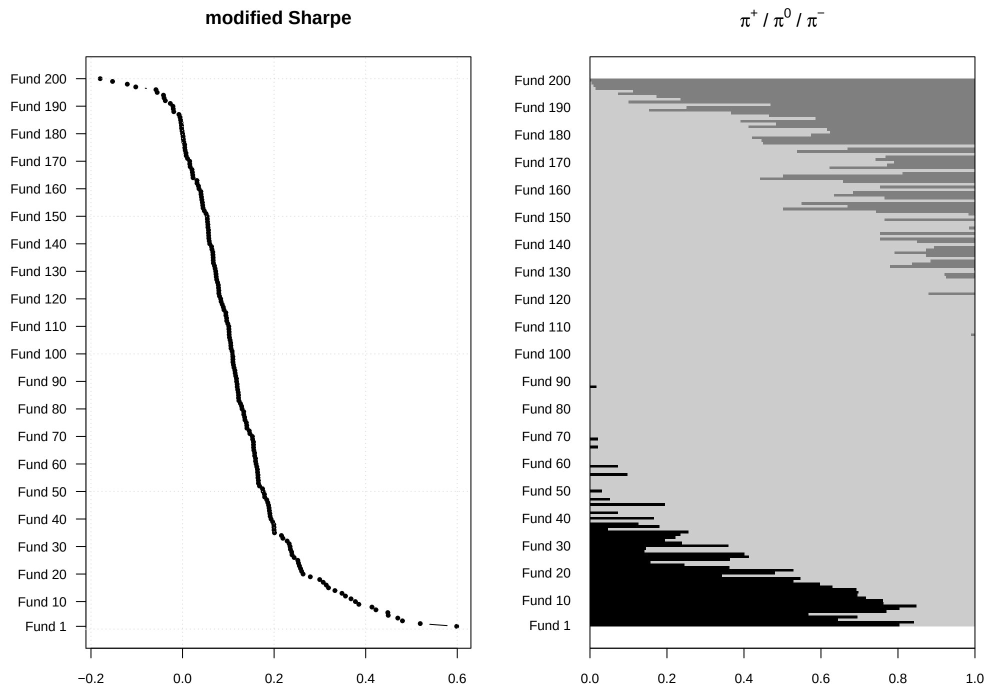


Figure 5: Left: Monthly modified Sharpe ratios for the two hundred funds in our database, ranked by decreasing modified Sharpe ratio. Right: Out-performance (black), equal performance (light gray) and under-performance (dark gray) ratios corresponding to the funds. The equal performance ratio is obtained with the block-bootstrap test with $B = 500$ and block-size set to four. The out- and under-performance ratios are obtained via the attributions (9) and (10).

Fund	Strategy	Overall	Equity Hedge	Event-Driven	Relative Value	Macro
AA	Relative Value	1 [0.81;0.00]	1 [1.00;0.00]	1 [0.75;0.00]	1 [0.77;0.00]	1 [0.77;0.00]
AB	Macro	2 [0.84;0.00]	1 [1.00;0.00]	1 [0.78;0.00]	2 [0.70;0.00]	1 [0.94;0.00]
AC	Relative Value	3 [0.64;0.00]	1 [1.00;0.00]	1 [0.66;0.00]	2 [0.67;0.00]	2 [0.52;0.00]
AD	Relative Value	4 [0.70;0.00]	1 [1.00;0.00]	1 [0.57;0.00]	3 [0.57;0.00]	2 [0.70;0.00]
AE	Relative Value	5 [0.57;0.00]	1 [1.00;0.00]	1 [0.57;0.00]	4 [0.66;0.00]	2 [0.22;0.00]
AF	Relative Value	6 [0.77;0.00]	1 [1.00;0.00]	1 [0.74;0.00]	5 [0.68;0.00]	2 [0.79;0.00]
AG	Macro	7 [0.80;0.00]	1 [1.00;0.00]	1 [0.78;0.00]	6 [0.61;0.00]	2 [0.93;0.00]
AH	Relative Value	8 [0.85;0.00]	1 [1.00;0.00]	1 [0.87;0.00]	6 [0.63;0.00]	3 [0.84;0.00]
AI	Event-Driven	9 [0.76;0.00]	1 [1.00;0.00]	1 [1.00;0.00]	7 [0.44;0.00]	3 [0.68;0.00]
AJ	Event-Driven	10 [0.76;0.00]	1 [1.00;0.00]	2 [0.98;0.02]	7 [0.50;0.00]	3 [0.68;0.00]
HI	Macro	191 [0.00;0.53]	47 [0.00;0.09]	48 [0.00;0.55]	50 [0.00;0.78]	49 [0.00;0.86]
HJ	Relative Value	192 [0.00;0.90]	47 [0.00;0.85]	48 [0.00;0.80]	50 [0.00;1.00]	50 [0.00;0.93]
HK	Equity Hedge	193 [0.00;0.78]	47 [0.00;0.66]	48 [0.00;0.71]	51 [0.00;0.88]	50 [0.00;0.87]
HL	Event-Driven	194 [0.00;0.83]	48 [0.00;0.63]	48 [0.00;0.87]	51 [0.00;0.89]	50 [0.00;0.95]
HM	Equity Hedge	195 [0.00;0.93]	48 [0.00;0.92]	49 [0.00;0.92]	51 [0.00;1.00]	50 [0.02;0.98]
HN	Equity Hedge	196 [0.00;0.89]	49 [0.00;0.83]	49 [0.01;0.92]	51 [0.00;0.95]	50 [0.00;0.95]
HO	Macro	197 [0.02;0.98]	50 [0.02;0.98]	49 [0.04;0.96]	51 [0.00;1.00]	50 [0.00;1.00]
HP	Event-Driven	198 [0.01;0.99]	50 [0.02;0.98]	49 [0.02;0.98]	51 [0.00;1.00]	51 [0.00;1.00]
HQ	Equity Hedge	199 [0.01;0.99]	50 [0.00;1.00]	50 [0.02;0.98]	51 [0.00;1.00]	51 [0.00;1.00]
HR	Event-Driven	200 [0.00;1.00]	51 [0.00;1.00]	50 [0.00;1.00]	51 [0.00;1.00]	51 [0.00;1.00]

Table 1: Fund vs peer universe. This table reports for the ten best and ten worst funds (in terms of modified Sharpe), the ranking with respect to a given universe: overall (200 funds), Equity-Hedge (50 funds), Event-Driven (50 funds), Relative Value (50 funds) and Macro (50 funds). The first number is the rank of the fund within the peer universe. The square parenthesis reports the out-performance and under-performance ratios, respectively. The ratios are obtained with the attribution rule in (9) and (10). The equal performance ratio is obtained with the block-bootstrap approach with block-size set to four.

	Equity Hedge	Event-Driven	Relative Value	Macro	Diff Strat
Equity Hedge	[0.10;0.17]	[0.03;0.32]	[0.01;0.31]	[0.00;0.25]	[0.01;0.26]
Event-Driven	[0.27;0.11]	[0.19;0.24]	[0.07;0.21]	[0.05;0.17]	[0.10;0.13]
Relative Value	[0.32;0.04]	[0.17;0.13]	[0.17;0.15]	[0.13;0.07]	[0.17;0.07]
Macro	[0.21;0.02]	[0.09;0.04]	[0.08;0.04]	[0.10;0.07]	[0.11;0.04]

Table 2: Strategy vs. strategy (peers vs. peers) analysis. The square parenthesis reports the average out-performance and under-performance ratios, respectively. The ratios are obtained with the attribution rule in (9) and (10). The equal performance ratio is obtained with the block-bootstrap approach with block-size set to four. Diff Strat aggregates the peer performance ratios computed using funds belonging to a different investment style.

7. Conclusion

The false discovery rate (FDR) approach of Storey (2002) is known to be a powerful tool to control the size in multiple testing. It has been recently applied by Barras et al. (2010) to estimate the proportion of truly out-performing mutual funds and by Bajgrowicz and Scaillet (2012) for the selection of out-performing technical trading rules. As we show in this paper, the FDR approach has also important applications in the analysis of peer performance of (hedge) funds as it allows to characterize the peer performance of a fund accounting for the estimation uncertainty in the performance measure used. We develop the estimator and analyze its properties in an extensive simulation study. In the empirical application we provide example graphs and tables that show how the proposed measures can be used as a screening tool in practice. The proposed tool is implemented in the open source (R Development Core Team, 2011) package **CompStrat** (Ardia and Boudt, 2013), soon available from the authors' website.

References

Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D., Monahan, J., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 60, 953–966.

Ardia, D., Boudt, K., 2013. CompStrat: Comparing Investment Strategies in R. Version 1-00.13.

Bajgrowicz, P., Scaillet, O., 2012. Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics 106, 473–491.

Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. Journal of Finance 65, 179–216.

Boudt, K., Peterson, B. G., Croux, C., 2008. Estimation and decomposition of downside risk for portfolios with non-normal returns. Journal of Risk, Winter 2008, 79–103.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Cornish, E. A., Fisher, R. A., 1937. Moments and cumulants in the specification of distributions. Revue de l'Institut International de Statistique 5 (4), 307–320.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 7, 1–26.

Favre, L., Galeano, J., 2002. Mean-modified value-at-risk with hedge funds. The Journal of Alternative Investments 5 (2), 21–25.

Fernández, C., Steel, M. F., 1998. On bayesian modelling of fat tails and skewness. Journal of the American Statistical Association 93 (441), 359–371.

Fung, W., Hsieh, D., 2004. Hedge fund benchmarks: A risk based approach. Financial Analysts Journal 60, 65–80.

Gregoriou, G. N. ., Gueyie, J.-P., 2003. Risk-adjusted performance of funds of hedge funds using a modified sharpe ratio. Journal of Wealth Management 6, 77–83.

Guidotti, I., Nagy, I., September 2011. Rational behavior of hedge fund managers and investors.

Hoberg, G., Kumar, N., Prabhala, N., 2013. Customized mutual fund peers. Working paper.

Lambert, P., Laurent, S., 2001. Modelling financial time series using GARCH-type models and a skewed Student density. Working paper, université de Liège.

Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the Sharpe ratio. Journal of Empirical Finance 15, 850–859.

Politis, D., Romano, J., 1992. Exploring the limits of bootstrap. John Wiley, New York, Ch. A circular block-resampling procedure for stationary data, pp. 263–270.

Politis, D., White, H., 2004. Automatic block-length selection for the dependent bootstrap. Econometric Reviews 23 (1), 53–70.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Scott, R. C., Horvath, P. A., September 1980. On the direction of preference for moments of higher order than the variance. Journal of Finance 35 (4), 915–919.

Sharpe, W. F., 1994. The Sharpe ratio. Journal of Portfolio Management Fall, 49–58.

Sortino, F. A., Price, L. N., 1994. Performance measurement in a downside risk framework. Journal of Investing Fall, 59–65.

Storey, J., 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society B 64, 479–498.

Tierney, L., Rossini, A. J., Li, N., Sevcikova, H., 2012. snow: Support for Simple Parallel Computing in R. Version 0.3-9-1.

Treynor, J. L., Black, F., 1973. How to use security analysis to improve portfolio selection. Journal of Business 46 (1), 66–86.

Zangari, P., 1996. A VaR methodology for portfolios that include options. RiskMetrics Monitor First Quarter, 4–12.