



## The peer performance ratios of hedge funds



David Ardia<sup>a,b,\*</sup>, Kris Boudt<sup>c,d</sup>

<sup>a</sup> Institute of Financial Analysis, University of Neuchâtel, Neuchâtel, Switzerland

<sup>b</sup> Department of Finance, Insurance and Real Estate, Laval University, Québec City, Canada

<sup>c</sup> Solvay Business School, Vrije Universiteit Brussel, Belgium

<sup>d</sup> School of Business and Economics, Vrije Universiteit Amsterdam, The Netherlands

### ARTICLE INFO

#### Article history:

Received 16 September 2016

Accepted 24 October 2017

Available online 26 October 2017

#### Keywords:

False discoveries

Hedge fund

Multiple hypothesis testing

Peer performance

Performance measurement

### ABSTRACT

We define the outperformance (resp. underperformance) of an investment fund as the percentage of funds in the peer universe for which the *true* performance of the focal fund is higher (resp. lower). We show that the *p*-values of the pairwise tests of equal performance can be used to obtain estimates of the out- and underperformance ratio that are robust to false discoveries – estimated alpha differentials for which the significance test has a low *p*-value while the true alpha is identical. When applied to hedge funds, we find that ranking funds on the outperformance ratio leads to a top quintile portfolio with a higher absolute and risk-adjusted performance than when the estimated alpha is used.

© 2017 Elsevier B.V. All rights reserved.

### 1. Introduction

Why does the estimated alpha of an investment fund differ from its peers? If the alpha of the peer funds is truly identical, estimation error must be the driving factor in a frequentist approach.<sup>1</sup> In their pioneer study, Barras et al. (2010) show how to accurately estimate the portion of true positive alpha funds in the universe. Their estimation approach exploits the properties of the discovery rate approach of Storey (2002) to avoid the common pitfall of also associating positive alpha to zero-alpha funds with a significant estimated positive alpha. In peer performance evaluation, a similar problem arises when estimating the percentage of peer funds that are outperformed by the investment fund of interest. In fact, the traditional approach to peer performance evaluation is to rank funds based on their estimated alpha, and then use the percentile ranks to classify peer performance as outperformance or underperformance. By construction, this approach ignores the possibility that funds in the peer group can have the same alpha, and thus tends to overestimate the outperformance and underperformance. Moreover, in the extreme case where all funds truly have the same alpha, the percentile-rank rate of outperformance obtained using the estimated alpha is a random number between zero and one, and thus has no information value.

Since institutional and retail investors tend to increase their portfolio allocation to outperforming funds (see, e.g., Fung et al., 2008), accounting for the presence of equal alpha among funds in the peer group is of crucial importance to avoid false discoveries and inefficient capital allocation. In this paper, we recommend evaluating peer performance of a fund using three peer performance parameters, for which we propose a non-parametric estimator that controls for false discoveries. Under the proposed triple-layered peer performance evaluation framework, a fund can exhibit three types of peer performance with respect to a universe: (i) *equal-performance*: the percentage composition of the peer universe in terms of funds that perform equally as well as the focal fund,<sup>2</sup> (ii) *outperformance*: the percentage of peer funds that underperform the focal fund, and (iii) *underperformance*: the percentage of peer funds that outperform the focal fund. Throughout the paper, we denote these parameters as  $\pi_i^0$ ,  $\pi_i^+$  and  $\pi_i^-$ , respectively, where *i* is the index of the focal fund.

The estimation of the peer performance parameters is challenging. In fact, the traditional approach of counting the percentage of estimated positive and negative alpha-differentials does not test whether the differences are statistically significant. It implicitly assumes  $\pi_i^0$  to be zero and thus overestimates  $\pi_i^+$  and  $\pi_i^-$ . The alternative approach of counting the percentage of significant positive and negative alpha-differentials relies on the estimated standard errors and the marginal distribution of *t*-statistics to account for the estimation error in a single pairwise test of equal performance.

<sup>2</sup> In the following, we define the focal fund as the fund of interest, for which the peer performance is measured.

\* Corresponding author.

E-mail addresses: [david.ardia@unine.ch](mailto:david.ardia@unine.ch) (D. Ardia), [kris.boudt@vub.be](mailto:kris.boudt@vub.be) (K. Boudt).

<sup>1</sup> We refer to Pástor and Stambaugh (2002) for the alternative Bayesian approach where the peer performance of a fund corresponds to an analysis of the credible set associated with the posterior distribution of the fund's alpha.

It however does not control for the false positives in the multiple hypothesis setup of testing the difference between the focal fund's alpha against all other peer funds. In practice, because of the small sample size in fund performance evaluation, there is also a problem of false negatives due to the lack of power of the  $t$ -test.

The solution that we provide in this paper is designed to account for these false discoveries. The proposed estimators of the peer performance parameters, called *peer performance ratios*, are obtained in two steps. We first take a frequentist approach to compute, for each of the  $n$  peer funds, the  $p$ -value of the null hypothesis that the alpha of the peer fund equals the alpha of the fund considered. False discoveries occur when, in the resulting sample of  $p$ -values, there are small  $p$ -values for funds with equal alpha. As shown by Storey (2002), Barras et al. (2010) and Bajgrowicz and Scaillet (2012), we can then exploit the distribution of  $p$ -values under the null hypothesis to obtain a reliable estimate of the peer performance ratios that is robust to the presence of false discoveries in the first-stage sample of  $p$ -values.

Alternatively, one could achieve robustness to false discoveries using the bootstrap techniques of Kosowski et al. (2005), Fama and French (2010), and Ferson and Chen (2015), or the parametric approach of Chen et al. (2017). The bootstrap requires comparing the obtained alpha-differential with artificially generated data samples where the variation in fund performance is due entirely to sample variability. Under the parametric setup of Chen et al. (2017), peer performance ratios could be obtained by modeling the alpha-differential as a realization from a mixture of normal distributions. Both the bootstrap and parametric approaches have the disadvantage of being computationally demanding. In contrast, our non-parametric approach has an explicit form and is thus simple to implement.

We perform an empirical validation of the proposed peer performance ratios on the union of all active and dead hedge funds in the Hedge Fund Research (HFR) database as of July 2014. The performance is evaluated based on the funds' monthly net returns. The peer category is defined as the group of hedge funds pursuing the same investment style (e.g., Macro funds). We account for the time-variation in the distribution of the hedge funds' alphas by calculating the peer performance ratios on monthly net returns observed over five-year rolling samples for a period ranging from January 2000 to June 2014. The resulting outperformance ratios are thus dynamic. We find that, for the majority of hedge funds, there is a large proportion of funds with equal performance. Therefore, on average, the standard approach for estimating peer performance using percentile ranks overestimates the outperformance and underperformance. For our sample of hedge funds, the average difference between the percentile-rank approach and the proposed estimator is around 33 percentage points.

Based on an extensive out-of-sample analysis, we find that robustness to false positives matters when predicting fund performance: The peer performance ratios have predictive value for future fund performance, and this predictive value is incremental to the information contained in alternative performance measures. Using portfolio sorts, we first show that compared to the traditional use of the fund's alpha, the use of the outperformance ratio to select the top quintile performing funds improves significantly the performance of the quarterly, semi-annually, and annually rebalanced portfolios. Second, we combine the various predictors of fund performance as regressors to forecast the hedge fund (risk-adjusted excess) return. These predictive regressions indicate that selecting funds using the outperformance ratio yields a portfolio with both a higher return and a higher risk-adjusted excess return. Importantly, this result is robust to adding other performance measures as predictors, as well as controlling for a host of other influences.

As an additional test of the validity of the proposed peer performance ratios, we use a simulation analysis to document their good finite sample bias properties. We also confirm standard predictions about the fund size and fund age on the cross-sectional distribution of peer performance. In particular, we find that when a fund has more assets under management, the outperformance ratio tends to be lower and, for two funds with the same age, the underperformance ratio is higher.

Altogether, the empirical results strongly support our view that, within a peer group, there is often an important proportion of hedge funds that have the same alpha. Ignoring this fact leads to false discoveries in terms of overestimating both the percentage of peers that the fund outperforms and that it underperforms. The proposed peer performance ratios are designed to avoid this problem. Their application to real-world hedge fund return data confirms their validity in terms of improved fund selection and estimation results that are aligned with model-based predictions. In practice, we recommend their use for *ex-ante* screening of the universe to select potential investments for a deeper, more fundamental, analysis of the fund's investment style, portfolio holdings, expenses, and organization, as well as for *ex-post* performance evaluation.

The remainder of this article is organized as follows. Section 2 introduces the methodology employed to obtain the peer performance ratios. Section 3 presents our empirical results regarding the importance of avoiding false discoveries in peer performance analysis for hedge funds.<sup>3</sup> In Section 4, we give our conclusions.

## 2. Estimation of peer performance ratios

In this study, we measure the peer performance of a focal fund  $i$  belonging to a peer universe of  $n + 1$  funds using three peer performance parameters: (i)  $\pi_i^0$ : the proportion of funds in the peer group that perform equally well as fund  $i$ , (ii)  $\pi_i^+$ : the proportion of funds in the peer group that are outperformed by fund  $i$ , and (iii)  $\pi_i^-$ : the proportion of funds in the peer group that outperform fund  $i$ . Two major strengths of the proposed peer performance ratios are that they require the relative performance between two funds to cross a threshold of statistical significance to be counted as evidence of a difference in performance, and that the false discovery rate methodology is used to obtain peer performance estimates that are robust to false positives.

### 2.1. Definitions

We consider a universe with a total of  $n + 1$  funds. We denote  $\Delta_{i-j}$  as the (true) difference in performance between fund  $i$  and  $j$  ( $i \neq j$ ), and  $\hat{\Delta}_{i-j}$  as the corresponding estimate. Throughout this study, we use the *hat* superscript symbol to denote sample-based estimates. Our objective is to estimate the percentage of funds that have equal, lower, or greater risk-adjusted performance than fund  $i$ . We denote by  $\pi_i^0$  ( $n_i^0$ ),  $\pi_i^+$  ( $n_i^+$ ), and  $\pi_i^-$  ( $n_i^-$ ), the proportion (number) of funds for which  $\Delta_{i-j} = 0$ ,  $\Delta_{i-j} > 0$ , and  $\Delta_{i-j} < 0$ , respectively.

The risk-adjusted performance of a fund is typically estimated by the intercept of the least squares regression of the fund returns on a series of risk factors, such as the four Carhart factors or the

<sup>3</sup> All of the computations performed in this study employed the R statistical computing language (R Development Core Team, 2017) with the package **PeerPerformance** (Ardia and Boudt, 2017), which is freely available at <https://CRAN.R-project.org/package=PeerPerformance>.

seven Fung and Hsieh hedge fund risk factors (Carhart, 1997; Fung and Hsieh, 2004).<sup>4</sup> Let  $\mathbf{f}_t$  be the  $(K \times 1)$  vector of risk factors at time  $t$ , and denote by  $r_{i,t}$  the fund's  $i$  return at time  $t$ . A test for equal performance of two funds  $i$  and  $j$  is obtained by testing the significance of the estimated intercept of the ordinary least squares (OLS) regression of  $(r_{i,t} - r_{j,t})$  on  $\mathbf{f}_t$ :

$$(r_{i,t} - r_{j,t}) = \Delta_{i-j} + \beta'_{i-j} \mathbf{f}_t + \varepsilon_{i-j,t}, \tag{1}$$

where  $\beta_{i-j}$  is the  $(K \times 1)$  vector of factor exposures and  $\varepsilon_{i-j,t}$  is the corresponding error term, for  $t = 1, \dots, T$ . The estimated intercept is:

$$\widehat{\Delta}_{i-j} = \frac{1}{T} \left( \sum_{t=1}^T (r_{i,t} - r_{j,t}) - \widehat{\beta}'_{i-j} \mathbf{f}_t \right),$$

where  $\widehat{\beta}_{i-j}$  is the least squares estimate of  $\beta_{i-j}$ . From the central limit theorem, it follows that under regularity conditions,  $\widehat{\Delta}_{i-j}$  is asymptotically normally distributed around  $\Delta_{i-j}$ . In our application, we compute its standard error  $\widehat{se}_{i-j}$  using the heteroscedasticity and autocorrelation robust (HAC) standard error estimators of Andrews (1991) and Andrews and Monahan (1992).

We denote  $\widehat{\tau}_{i-j} \equiv \widehat{\Delta}_{i-j} / \widehat{se}_{i-j}$  as the studentized test statistic such that when the absolute value of  $\widehat{\tau}_{i-j}$  is higher, the evidence against the  $H_0$  of equal performance is greater. The  $p$ -values are then defined as two times the (estimated) probability integral transform of minus the absolute value of  $\widehat{\tau}_{i-j}$  under  $H_0$ :

$$\widehat{p}_{i-j} \equiv 2\widehat{F}_{i-j}(-|\widehat{\tau}_{i-j}|),$$

where  $\widehat{F}_{i-j}$  is a consistent estimate of the true cumulative distribution function  $F_{i-j}$  of  $\widehat{\tau}_{i-j}$  under  $H_0$ . In our application, we set  $\widehat{F}_{i-j}$  to the standard normal cumulative distribution.<sup>5</sup>

### 2.2. Existing estimators of peer performance

Several estimators for the peer performance parameters  $\pi_i^+$ ,  $\pi_i^0$ , and  $\pi_i^-$  can be considered. First, let us look at the percentile-rank estimation approach, which estimates the outperformance ratio  $\pi_i^+$  as the percentage of funds for which the estimated performance of the focal fund is higher:

$$\widehat{\Delta}_i^+ \equiv \frac{1}{n} \sum_{j \neq i} I\{\widehat{\Delta}_{i-j} \geq 0\}, \tag{2}$$

where  $I\{\cdot\}$  denotes the indicator function, which equals one if the condition holds and zero otherwise. Similarly,  $\widehat{\Delta}_i^-$  is given by the percentage of funds for which  $\widehat{\Delta}_{i-j} < 0$ . Clearly, the percentile-rank approach has the disadvantage of implicitly setting the percentage of equal performance to zero ( $\widehat{\Delta}_i^0 = 0$ , for all  $i$ ), and thus overestimating the percentage of out- and underperformance.

A second simple but biased estimator of the outperformance ratio  $\pi_i^+$  is the percentage of funds for which the estimated test statistic  $\widehat{\tau}_{i-j}$  exceeds the (estimated)  $\gamma^+$ -quantile of the distribution of  $\widehat{\tau}_{i-j}$  under the null hypothesis, which is denoted by  $\widehat{q}_{i-j}^{\gamma^+}$ . Typical values for  $\gamma^+$  are 90% and 95%. The corresponding estimator is then:

$$\widehat{\tau}_i^+ \equiv \frac{1}{n} \sum_{j \neq i} I\{\widehat{\tau}_{i-j} \geq \widehat{q}_{i-j}^{\gamma^+}\}. \tag{3}$$

<sup>4</sup> For presentation purposes, we focus on the fund's alpha as the risk-adjusted performance measure. However, it is straightforward to apply the proposed peer performance evaluation framework with other risk-adjusted performance measures, such as the fund's (modified) Sharpe ratio, by using the equal-performance test of Ledoit and Wolf (2008) and Ardia and Boudt (2015).

<sup>5</sup> We considered alternative estimators of  $F_{i-j}$  based on (block) bootstrapping the empirical distribution, but the results are qualitatively similar and thus we omit them here for brevity.

A left-sided threshold  $\widehat{q}_{i-j}^{\gamma^-}$  is used to compute the underperformance ratio as the percentage of funds for which  $\widehat{\tau}_{i-j} \leq \widehat{q}_{i-j}^{\gamma^-}$ , where  $\gamma^-$  is typically 5% or 10%. The percentage of equal performance is estimated as the percentage of funds for which the  $t$ -statistic is between  $\widehat{q}_{i-j}^{\gamma^-}$  and  $\widehat{q}_{i-j}^{\gamma^+}$ :

$$\widehat{\tau}_i^0 \equiv \frac{1}{n} \sum_{j \neq i} I\{\widehat{\tau}_{i-j} > \widehat{q}_{i-j}^{\gamma^-}\} \cdot I\{\widehat{\tau}_{i-j} < \widehat{q}_{i-j}^{\gamma^+}\}. \tag{4}$$

Both the outperformance ratio estimates obtained using the percentile-rank and pairwise significance tests are characterized by a potentially large number of false positives (i.e., detected instances of outperformance when in fact there is equal-performance or underperformance) and false negatives (i.e., detected equal- or underperformance when in fact there is outperformance). Formally, we have:

$$\begin{aligned} \sum_{j \neq i} I\{\widehat{\Delta}_{i-j} \geq 0\} &= \underbrace{\sum_{\substack{j \neq i \\ \Delta_{i-j} > 0}} I\{\widehat{\Delta}_{i-j} \geq 0\}}_{\text{correct classifications}} \\ &+ \underbrace{\sum_{\substack{j \neq i \\ \Delta_{i-j} = 0}} I\{\widehat{\Delta}_{i-j} \geq 0\} + \sum_{\substack{j \neq i \\ \Delta_{i-j} < 0}} I\{\widehat{\Delta}_{i-j} \geq 0\}}_{\approx n_i^0/2} \\ &\quad \underbrace{\hspace{10em}}_{\text{false positives}} \\ &- \underbrace{\sum_{\substack{j \neq i \\ \Delta_{i-j} > 0}} I\{\widehat{\Delta}_{i-j} < 0\}}_{\text{false negatives}}, \end{aligned} \tag{5}$$

and:

$$\begin{aligned} \sum_{j \neq i} I\{\widehat{\tau}_{i-j} \geq \widehat{q}_{i-j}^{\gamma^+}\} &= \underbrace{\sum_{\substack{j \neq i \\ \Delta_{i-j} > 0}} I\{\widehat{\tau}_{i-j} \geq \widehat{q}_{i-j}^{\gamma^+}\}}_{\text{correct classifications}} \\ &+ \underbrace{\sum_{\substack{j \neq i \\ \Delta_{i-j} = 0}} I\{\widehat{\tau}_{i-j} \geq \widehat{q}_{i-j}^{\gamma^+}\} + \sum_{\substack{j \neq i \\ \Delta_{i-j} < 0}} I\{\widehat{\tau}_{i-j} \geq \widehat{q}_{i-j}^{\gamma^+}\}}_{\approx n_i^0(1-\gamma^+)} \\ &\quad \underbrace{\hspace{10em}}_{\text{false positives}} \\ &- \underbrace{\sum_{\substack{j \neq i \\ \Delta_{i-j} > 0}} I\{\widehat{\tau}_{i-j} < \widehat{q}_{i-j}^{\gamma^+}\}}_{\text{false negatives}}. \end{aligned} \tag{6}$$

For the percentile-rank approach in (2) and (5), we expect that for most funds, the number of false positives exceeds the number of false negatives, and thus that the estimated outperformance is overestimated. In case of the significance testing approach in (3) and (6), the balance depends on the underlying return generating process and the probability level  $\gamma^+$  used for testing. As can be seen in (6), we have that, for lower values of the critical value  $\widehat{q}_{i-j}^{\gamma^+}$ , there are more false positives, and less false negatives. This approach accounts for the estimation error in the setting of pairwise testing, but not the joint hypothesis of testing equal-performance with the peer funds.

We thus need an estimator for the outperformance ratio that is robust to false positives under a framework of multiple hypothesis testing. As a solution, it might be tempting to use a sequence of multiple hypothesis tests, such as Hotelling's  $T^2$  test of equality of Sharpe ratios, with family-wise error rate control of Type I

errors over the sequence of tests.<sup>6</sup> The disadvantage of those multiple tests is that their applicability to our problem is reduced when the number of peer funds is large. The standard setup requires that the time series is available for the same period for all funds, which limits the number of observations to the time span of the fund with the shortest history.

2.3. Proposed estimation strategy

The estimator we propose combines the advantages of pairwise and multiple testing in a two-step estimation procedure. For each fund, we first estimate the percentage of peer funds with equal performance in an unbiased manner using only pairwise tests of equal performance between the focal fund and a peer fund. After performing this procedure for each potential pair, a sample of  $p$ -values  $\hat{p}_{i-j}$  associated with a two-sided test of the null hypothesis  $H_0 : \Delta_{i-j} = 0$ , for  $j = 1, \dots, n, j \neq i$ , is obtained. For fixed  $i$ , the distribution of the  $p$ -values is (asymptotically) a mixture of  $p$ -values that are uniformly distributed (the pairs for which the null hypothesis is correct) and  $p$ -values that are close to zero (when the null hypothesis is false). This key insight is provided by Storey (2002) and it is used by Barras et al. (2010) to estimate the proportion of funds that perform equally well in the same manner as a passive investor in style indices.<sup>7</sup>

The next step is then to attribute the remaining segment of the peer group to funds that significantly underperform and those that outperform. Thus, for each fund, we obtain an *equal-performance ratio*, an *outperformance ratio*, and an *underperformance ratio*. The proposed equal-performance ratio is robust to false discoveries, and unless the fund performance differs significantly from its peers, it will tend toward 100%. When the return data are sufficiently informative about differences in performance, the outperformance ratio is suitable for classifying funds into top-performing funds. The estimators rely on pairwise tests to calculate the  $p$ -values, thus they can use the longest common time series span for each pair in an optimal manner, and parallel computing can be employed to calculate these  $p$ -values in a numerically efficient way.

2.4. The equal-performance ratio

A crucial feature of the proposed peer performance ratios is that they exploit the difference in the distribution of the  $p$ -values  $\hat{p}_{i-j}$  when  $\Delta_{i-j} = 0$  versus  $\Delta_{i-j} \neq 0$ .

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

$$(A1) : \mathbb{P}[\hat{p}_{i-j} < \lambda_i \mid \Delta_{i-j} \neq 0] = 1. \tag{7}$$

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

<sup>6</sup> This sequential testing approach involves first testing the null hypothesis of equal performance of the focal fund and the  $n$  peer funds. If the hypothesis is rejected, an elimination rule (e.g., removing the fund that contributes most to the test statistic) is then used subsequently to test for equal performance with the  $n - 1$  remaining peer funds. This procedure is then repeated until the null hypothesis is no longer rejected. This test is similar to the *reality check test* of White (2000), the *superior predictive ability test* of Hansen (2005), and the *model confidence set* of Hansen et al. (2011).

<sup>7</sup> Instead of estimating the market-wide equal-performance ratio, as in Barras et al. (2010), we estimate the individual equal-performance ratio for each fund, which broadens the application scope. The aggregate equal-performance ratio allows us to answer general economic questions such as the usefulness of actively managed funds, but our proposed individual equal-performance ratio can be used directly by investors to evaluate the performance of a specific fund.

In the case of equal performance, and provided that the estimated  $\hat{F}_{i-j}$  coincides with the true  $F_{i-j}$  used to calculate the  $p$ -value, the  $p$ -value is uniformly distributed for a given pair  $(i, j)$ . This implies that the probability of  $\hat{p}_{i-j}$  exceeding  $\lambda_i$  when  $\Delta_{i-j} = 0$  is  $1 - \lambda_i$ :

$$(A2) : \mathbb{P}[\hat{p}_{i-j} \geq \lambda_i \mid \Delta_{i-j} = 0] = 1 - \lambda_i. \tag{8}$$

In practice, the cumulative distribution function  $F_{i-j}$  is not fully known and the calculation of the  $p$ -values requires parameter estimates. Asymptotically, the  $p$ -value is uniformly distributed if consistent estimators are used (Rosenblatt, 1952), whereas in finite samples, assumption (A2) is only approximately satisfied.

A key result related to the definition of the proposed equal-performance ratio is that under (A1) and (A2), the expected number of  $p$ -values exceeding  $\lambda_i$  is  $(1 - \lambda_i)n_i^0$ , with  $n_i^0$  the number of peer funds that perform equally well as the focal fund:

$$\begin{aligned} \mathbb{E} \left[ \sum_{j \neq i} I\{\hat{p}_{i-j} \geq \lambda_i\} \right] &= \mathbb{E} \left[ \underbrace{\sum_{\substack{j \neq i \\ \Delta_{i-j} = 0}} I\{\hat{p}_{i-j} \geq \lambda_i\}}_{=(1-\lambda_i)n_i^0} \right] \\ &+ \mathbb{E} \left[ \underbrace{\sum_{\substack{j \neq i \\ \Delta_{i-j} \neq 0}} I\{\hat{p}_{i-j} \geq \lambda_i\}}_{=0} \right] = (1 - \lambda_i)n_i^0. \end{aligned}$$

Hence, a natural estimator for  $n_i^0$  is the number of estimated  $p$ -values exceeding  $\lambda_i$  divided by  $1 - \lambda_i$ :

$$\hat{n}_i^0 \equiv c_i^0 \min \left\{ \frac{\sum_{j \neq i} I\{\hat{p}_{i-j} \geq \lambda_i\}}{1 - \lambda_i}, n \right\}, \tag{9}$$

where we include an additional adjustment to bound from above the extrapolation to the number of peer funds, and  $c_i^0$  is a correction factor that adjusts for the bias induced by the truncation; see Appendix A. The corresponding estimate for the proportion of equal performance is:

$$\hat{\pi}_i^0 \equiv \frac{\hat{n}_i^0}{n}. \tag{10}$$

Since the estimator  $\hat{\pi}_i^0$  has the form of a sample average, we expect that in most relevant cases, the estimator  $\hat{\pi}_i^0$  is not only unbiased but also consistent for  $\pi_i^0$ . The proof of this requires a suitable law of large numbers that allows for the dependence in the  $p$ -values.<sup>8</sup>

The practical computation of the equal-performance ratio thus requires us to choose the threshold value  $\lambda_i$ . The larger the value of  $\lambda_i$ , the more likely it is that assumptions (A1) and (A2) are satisfied, but the fewer observations enter the summation used to estimate the equal-performance ratio in (9)–(10). We address this bias-variance trade-off in the estimation of  $\pi_i^0$  by optimizing the choice of  $\lambda_i$  based on a data-driven approach to determine the value of  $\lambda_i \in \{0.3, 0.32, \dots, 0.7\}$  which minimizes the estimated mean squared estimation error of  $\hat{\pi}_i^0$ . See Appendix B for more details. In our application to hedge funds,  $\lambda_i$  varies between

<sup>8</sup> Each  $p$ -value is uniformly distributed under the null hypothesis, but because of the correlation across hedge funds and the comparison with a common peer, the  $p$ -values corresponding to the different  $\Delta_{i-j} = 0$  tests are not uniformly distributed for different  $i$  and  $j$ . In the case of excessively high dependence, this may make the estimator inconsistent (e.g., when all  $p$ -values are identical). This strong dependence can occur in the rare case where a fund outperforms its peers by large amounts, where the data are not sufficiently informative to distinguish between a zero-alpha fund and a talented fund. See Bajgrowicz and Scaillet (2012, Appendix F) and the references therein.

0.3 and 0.7 with a median (resp. mean) value of 0.64 (0.56). In most cases, the concern of avoiding bias thus dominates the objective of precision in estimating the equal–performance ratios.

2.5. The out- and underperformance ratios

Given the estimate of the number of peer funds with equal performance  $\hat{n}_i^0$ , and the observed performance differences  $\hat{\Delta}_{i-j}$ , we then estimate the number of funds that are outperformed by the focal fund  $i$ ,  $n_i^+$ , and those that outperform fund  $i$ ,  $n_i^-$ . The attribution is based on the number of significant performance differences (using the studentized test statistic  $\hat{\tau}_{i-j} \equiv \hat{\Delta}_{i-j} / \widehat{SE}_{i-j}$ ) and an adjustment for false discoveries.

Clearly, given an estimate of  $n_i^+$ , we obtain an estimate for  $n_i^-$  by the requirement that  $n_i^+ + n_i^- = n - n_i^0$ , and vice versa. We start the estimation procedure on the side for which we have most observations. If there are more point estimates of outperformance by fund  $i$  (i.e., when  $\sum_{j \neq i} I\{\hat{\Delta}_{i-j} \geq 0\} \geq n/2$ ), we first estimate  $n_i^+$ , and then attribute  $n - \hat{n}_i^0 - \hat{n}_i^+$  to  $\hat{n}_i^-$ . The estimate  $\hat{n}_i^+$  is based on the fact that  $n_i^+$  corresponds to the naive estimate of the number of peer funds that are significantly outperformed by the focal fund (at a one-sided confidence level  $\gamma^+$ , that is, the number of funds for which the estimated test statistic  $\hat{\tau}_{i-j}$  exceeds the (estimated)  $\gamma^+$ -quantile of the distribution of  $\hat{\tau}_{i-j}$  under the null hypothesis, which is denoted by  $\hat{q}_{i-j}^{\gamma^+}$ ,<sup>9</sup> adjusted for the false positives (i.e., cases where  $\hat{\tau}_{i-j} \geq \hat{q}_{i-j}^{\gamma^+}$  when  $\Delta_{i-j} \leq 0$ ) and false negatives (i.e., cases where  $\hat{\tau}_{i-j} < \hat{q}_{i-j}^{\gamma^+}$  when  $\Delta_{i-j} > 0$ ):

$$\begin{aligned} n_i^+ &\equiv \sum_{j \neq i} I\{\Delta_{i-j} > 0\} \\ &= \sum_{j \neq i} I\{\hat{\tau}_{i-j} \geq \hat{q}_{i-j}^{\gamma^+}\} - \underbrace{\sum_{\substack{j \neq i \\ \Delta_{i-j}=0}} I\{\hat{\tau}_{i-j} \geq \hat{q}_{i-j}^{\gamma^+}\} - \sum_{\substack{j \neq i \\ \Delta_{i-j}<0}} I\{\hat{\tau}_{i-j} \geq \hat{q}_{i-j}^{\gamma^+}\}}_{\approx n_i^0(1-\gamma^+) \text{ false positives}} \\ &\quad + \underbrace{\sum_{\substack{j \neq i \\ \Delta_{i-j}>0}} I\{\hat{\tau}_{i-j} < \hat{q}_{i-j}^{\gamma^+}\}}_{\text{false negatives}} \\ &\approx \sum_{j \neq i} I\{\hat{\tau}_{i-j} \geq \hat{q}_{i-j}^{\gamma^+}\} - n_i^0(1 - \gamma^+). \end{aligned} \tag{11}$$

The approximation error in the last step of (11) is small for well-chosen values of  $\gamma^+$ . First, given  $\hat{n}_i^0$ , we can infer that the number of false positives is  $\hat{n}_i^0(1 - \gamma^+)$  if  $\Delta_{i-j} = 0$ . As  $\hat{n}_i^0$  is an unbiased estimate of  $n_i^0$ , so is  $\hat{n}_i^0(1 - \gamma^+)$  an unbiased estimate for the false discoveries when  $\Delta_{i-j} = 0$ . Since there are relatively more cases where  $\hat{\Delta}_{i-j}$  is positive rather than negative, the data indicate that the fund is more likely to be outperforming than underperforming. Therefore, the number of false positives is negligible when  $\Delta_{i-j} < 0$  and false negatives are avoided by setting  $\gamma^+$  as sufficiently low.<sup>10</sup>

<sup>9</sup> In our application, we use the asymptotic distribution of  $\hat{\tau}_{i-j}$  and thus we set  $\hat{q}_{i-j}^{\gamma^+}$  as the  $\gamma^+$ -quantile of the standard Normal distribution.

<sup>10</sup> For the estimation of the number of good and bad managers, Barras et al. (2010) follow the same reasoning. Ferson and Chen (2015) note that, when for many peer funds  $j$ , the difference in performance  $\Delta_{i-j}$  is not zero but small, there is a tendency to overestimate the equal–performance ratio. We do not worry much about the overestimation of  $\pi_i^0$  when  $\Delta_{i-j} = 0$ , because, economically speaking, the overestimation of  $\pi_i^0$  and the likely underestimation of both  $\pi_i^+$  and  $\pi_i^-$  is acceptable, as it reflects the simple fact that the data is not sufficiently informative to distinguish equal–performance from out- or

underperformance case where there are more point estimates of underperformance by fund  $i$  (i.e., when  $\sum_{j \neq i} I\{\hat{\Delta}_{i-j} \geq 0\} < n/2$ ), almost all of the peer funds either perform equally well or outperform fund  $i$ , such that  $n_i^-$  is accurately estimated by  $\sum_{j \neq i} I\{\hat{\tau}_{i-j} \leq \hat{q}_{i-j}^{\gamma^-}\} - \hat{n}_i^0\gamma^-$ , provided that we set  $\gamma^-$  as sufficiently high so that the number of false negatives becomes negligible. In the application, we set  $\gamma^+ = 0.4$  to avoid the false negatives and  $\gamma^- = 0.6$  to avoid the false positives.

Thus, by including an additional adjustment to the extrapolation, we obtain the following natural definitions of the outperformance and underperformance ratios of fund  $i$ :

$$\hat{\pi}_i^+ \equiv \begin{cases} \frac{1}{n} \max \left\{ \sum_{j \neq i} I\{\hat{\tau}_{i-j} \geq \hat{q}_{i-j}^{\gamma^+}\} - \hat{n}_i^0(1 - \gamma^+), 0 \right\} & \text{if } \sum_{j \neq i} I\{\hat{\Delta}_{i-j} \geq 0\} \geq n/2 \\ 1 - \hat{\pi}_i^0 - \hat{\pi}_i^- & \text{otherwise,} \end{cases}$$

and:

$$\hat{\pi}_i^- \equiv \begin{cases} \frac{1}{n} \max \left\{ \sum_{j \neq i} I\{\hat{\tau}_{i-j} \leq \hat{q}_{i-j}^{\gamma^-}\} - \hat{n}_i^0\gamma^-, 0 \right\} & \text{if } \sum_{j \neq i} I\{\hat{\Delta}_{i-j} \geq 0\} < n/2 \\ 1 - \hat{\pi}_i^0 - \hat{\pi}_i^+ & \text{otherwise.} \end{cases}$$

It is important to note that, compared to the approach of counting the significant alpha-differential estimates in (3), the outperformance and underperformance ratios explicitly correct for the presence of false positives by subtracting the terms  $\hat{n}_i^0(1 - \gamma^+)$  and  $\hat{n}_i^0\gamma^-$ , respectively.

2.6. Illustration on simulated return data

We now perform a Monte Carlo study to compare the finite sample bias of the proposed peer performance ratios with the alternative, more simple percentile-rank and significance test-based estimates discussed in Section 2.1. We present the results for estimating the outperformance ( $\pi^+$ ) and equal–performance ( $\pi^0$ ) parameters. The significance test-based estimates in (3) and (4) are implemented using either  $\gamma^+ = 90\%$  and  $\gamma^- = 10\%$ , or  $\gamma^+ = 95\%$  and  $\gamma^- = 5\%$ . We refer to the former using  $\hat{\tau}_{0.10}^+$  and  $\hat{\tau}_{0.10}^-$ , and use  $\hat{\tau}_{0.05}^+$  and  $\hat{\tau}_{0.05}^-$  for the latter.

The Monte Carlo setup is as follows. We assume a population of  $n = 250$  peer funds for which fund returns have the same volatility, but 10% (25 funds), 80% (200 funds) and 10% (25 funds) of all funds have an average annual return of 5%, 15% and 25%, respectively. For each replication, we draw  $T = 60$  monthly returns either from a multivariate Normal distribution or via a Normal copula-marginal approach with skewed Student- $t$  marginal distributions with fat tails and negative skewness.<sup>11</sup> The pairwise correlation is set to 50% in both cases. To evaluate the effect of overlap in the return distribution on the bias of the estimators, we consider three scenarios for the volatility  $\sigma$ . In the “Low” volatility scenario, we set the annual standard deviation to 5%. In the “Medium” volatility scenario, we set it to 7.5%. Finally, in the “High” volatility scenario, we set it to 10%. We perform 1000 replications and measure the bias of the equal- and outperformance ratios, computed as the average (over the Monte Carlo replications) of the differences between the estimator of equal- or outperformance ratio, and the true equal- or outperformance ratio. The true value of

underperformance. Moreover, in contrast with Barras et al. (2010) and Ferson and Chen (2015), the outcome of our analysis is not a statistic describing the whole universe, but a triplet of peer performance measures for each hedge fund, for which the ordinal interpretation is robust to potential tendencies to overestimate the equal–performance ratio in case the data is not sufficiently informative.

<sup>11</sup> We take the setup of Ardia and Boudt (2015) and rely on the standardized skewed Student- $t$  marginal distributions with degrees of freedom parameter  $\nu = 6$  and asymmetry parameter  $\xi = 0.75$ , which were calibrated on a universe of hedge fund returns similar to the one used in our empirical study.

**Table 1**

Monte Carlo results – Bias estimation.

This table presents the Monte Carlo results of the finite sample bias of the various estimators of the equal- and outperformance ratio parameters when the universe consists of  $n = 250$  peer funds. All funds have returns from the same distribution, except that 10% (25 funds), 80% (200 funds) and 10% (25 funds) of all funds have an average annual return of 5%, 15% and 25%, respectively. We perform 1000 replications for which we draw  $T = 60$  monthly returns from a multivariate Normal distribution (Panel A) or via a Normal copula–marginal approach (Panel B) with skewed Student- $t$  marginal distributions with fat tails and negative skewness. The pairwise correlation is set to 50% in both cases. We visualize the impact of overlapping distributions on the finite sample bias by considering a low ( $\sigma = 5\%$ ), medium ( $\sigma = 7.5\%$ ) and high volatility ( $\sigma = 10\%$ ) scenario. The peer-performance estimators considered are the peer-performance ratios:  $\hat{\pi}^0$  and  $\hat{\pi}^+$ , the percentile-rank estimators:  $\hat{\Delta}^0$  and  $\hat{\Delta}^+$ , and significance test-based estimates at the 10% and 5% levels:  $\hat{\tau}_{0.10}^0$ ,  $\hat{\tau}_{0.10}^+$ ,  $\hat{\tau}_{0.05}^0$  and  $\hat{\tau}_{0.05}^+$ . See Section 2.6 for details.

	Equal-performance				Outperformance			
	$\hat{\pi}^0$	$\hat{\Delta}^0$	$\hat{\tau}_{0.10}^0$	$\hat{\tau}_{0.05}^0$	$\hat{\pi}^+$	$\hat{\Delta}^+$	$\hat{\tau}_{0.10}^+$	$\hat{\tau}_{0.05}^+$
Panel A: Multivariate Normal with pairwise correlations of 50%								
$\sigma = 5\%$	-0.014	-0.660	-0.070	-0.035	0.007	0.330	0.035	0.017
$\sigma = 7.5\%$	-0.008	-0.660	-0.041	0.013	0.004	0.330	0.020	-0.006
$\sigma = 10\%$	0.021	-0.660	0.018	0.088	-0.010	0.330	-0.009	-0.044
Panel B: Copula-Skewed Student- $t$ marginals with pairwise correlations of 50%								
$\sigma = 5\%$	-0.017	-0.660	-0.069	-0.033	0.010	0.330	0.034	0.017
$\sigma = 7.5\%$	-0.009	-0.660	-0.038	0.016	0.006	0.330	0.019	-0.008
$\sigma = 10\%$	0.020	-0.660	0.020	0.090	-0.008	0.330	-0.010	-0.045

equal-performance ratio within the groups is 10%, 80% and 10%, respectively. The true value of outperformance ratio within the groups is 0%, 10% and 90%, respectively.

First, let us consider in Table 1 the bias of the percentile-rank approach to estimating the equal-performance parameter. Since  $\Delta^0 = 0$ , we have that its bias in all scenarios equals  $0.1 \times (0 - 0.1) + 0.8 \times (0 - 0.8) + 0.1 \times (0 - 0.1) = -0.66$ . Because of the underestimation of  $\pi^0$ , the outperformance and underperformance ratio are overestimated, resulting in a bias of 0.33 for  $\Delta^+$  in all scenarios considered. More interesting is the finite sample bias of the proposed equal-performance ratio  $\hat{\pi}^0$  versus the estimator based on computing the percentage of pairwise differences in performance at the 10% and 5% level, which are denoted by  $\hat{\tau}_{0.10}^0$  and  $\hat{\tau}_{0.05}^0$ . Parameter  $\hat{\pi}^0$  has a finite sample bias because of the small sample size used for testing ( $T = 60$ ,  $n = 250$ ), while  $\hat{\tau}_{0.10}^0$  and  $\hat{\tau}_{0.05}^0$  are intrinsically biased because of the false positives that arise in pairwise testing (as shown in (6)). For all three estimators, the finite sample bias increases (in absolute value) when the return distributions of the three types of funds have a higher overlap, that is, when  $\sigma$  is higher, and it thus becomes more difficult to distinguish equal-performance from outperformance and underperformance using a finite sample.

In our setup, the finite sample bias of  $\hat{\pi}^0$  is negligible. The finite sample bias of  $\hat{\tau}_{0.10}^0$  tends to be negative when there is little overlap, because of the large number of false positives when testing at the 10% level. There are fewer false positives when testing at the 5% level, at the expense of a higher number of false negatives. Indeed, the large positive bias of  $\hat{\tau}_{0.05}^0$  is due to the fact that, at a low significance level and thus a high cutoff value, many estimated differences in performance are not detected as significant. They thus inflate the estimation of the equal-performance ratio. In contrast, the proposed peer performance ratio uses a cutoff  $\lambda_i \in \{0.3, 0.32, \dots, 0.7\}$  and is therefore by design more robust to false negatives than when statistical significance tests are used at the 10% or 5% one-sided significance levels. Finally, note in the right part of Table 1, that the finite sample bias of the estimator of  $\pi^+$  is around minus one half of the bias of the corresponding estimator of  $\pi^0$ . The good finite sample bias properties of the equal-performance ratio  $\hat{\pi}^0$  thus also lead to good finite sample properties of  $\hat{\pi}^+$  (and  $\hat{\pi}^-$ , since  $\hat{\pi}^0 + \hat{\pi}^+ + \hat{\pi}^- = 1$ ).

### 3. Application to hedge funds

We now illustrate the practical relevance of the proposed peer performance ratios to analyze the performance of hedge funds. We

first introduce the empirical setup regarding data used and choices in implementation. We then quantify the overestimation of outperformance and underperformance when the traditional percentile rank-based outperformance is used. Next, we use a portfolio sorting analysis to assess the out-of-sample gains in performance that can be achieved by selecting funds on their outperformance ratio rather than on their estimated alpha. Finally, we verify that the cross-section of peer performance is consistent with theoretical predictions about the determinants of peer performance.

#### 3.1. Data description

Our universe consists of the active and dead U.S. hedge funds included in the Hedge Fund Research (HFR) database as of July 2014. To account for the time-variation in the distribution of the hedge funds' alpha, we use rolling five-year estimation windows of monthly hedge fund net returns and factor data over the period 2000–2014.<sup>12</sup> This thus leads to 39 estimation dates and a total of 15,370 funds. To avoid survivorship bias, the fund composition of each sample changes, and tracks the funds available in the HFR alive and dead funds databases for the corresponding period. On each of the (rolling) quarterly database updates, we exclude funds with less than 60 available observations, and keep U.S. funds pursuing either an Equity Hedge, Event-Driven, Relative Value or Macro investment style. We further delete incoherent and duplicate entries following the approach described in Joenväärä et al. (2016). To diminish the risk of backfill bias, we remove the first 12-month history of each hedge fund in our database, such that the effective number of observations per estimation window ranges between 48 and 60. On average, there are 1391 funds per (rolling) quarterly sample, with a minimum of 921 funds and a maximum of 1634 funds. Each sample starts at the beginning of the quarter and is indexed by  $q = 1, \dots, 39$  (i.e., January 2000–December 2004, April 2000–March 2005, ..., July 2009–June 2014).

We define the peer group as the set of hedge funds following the same investment style (Equity Hedge, Event-Driven, Macro, and Relative Value). As the pairwise peer performance measure, we use the alpha-differential obtained as the intercept in the linear factor model in (1), estimated using the nine risk factors constituting the union of the four factors in Carhart (1997) and the

<sup>12</sup> The choice of a five-year window length is motivated by the requirement to ensure a sufficient sample size to estimate the peer performance ratios with acceptable accuracy, while allowing the estimated peer performance ratios to be responsive to changes in the underlying hedge fund's performance.

**Table 2**

## Summary statistics.

This table presents the average value of the number of funds, the fund performance, and the fund characteristics over the quarterly updated five-year estimation samples between January 2000 and June 2014 (average over 39 samples). Panel A reports the summary statistics for the annualized performance (geometric return in percent) of each fund, averaged across all funds, the Equity Hedge, Event-Driven, Macro, and Relative Value hedge funds. N (rounded) is the number of funds, Mean is the average, Med is the median, Std is the standard deviation, Min/Max are the minimum and maximum values, and 25th/75th are the 25th and 75th percentiles. Panel B reports the funds' monthly nine-factor model alpha values (in percent, annualized). Panel C reports the distribution of the average value of the fund characteristics (as observed at the beginning of each sample). AUM is the asset under management expressed in million US dollars. Age is the fund's age in months since the fund's inception date. Management and performance fees are in percent. Leverage, hurdle rate, and high watermark are dummy variables. Note that because of the averaging per group, the average maximum value of all funds is not equal to the maximum of the average values per fund style. See Section 3.1 for details.

	N	Mean	Med	Std	Min	25th	75th	Max
Panel A: Fund compound returns (in percent, annualized)								
All Funds	1391	8.49	7.56	9.28	-35.25	3.58	12.27	75.98
Equity Hedge	715	8.51	7.50	9.67	-32.90	3.46	12.46	71.03
Event-Driven	163	8.61	7.79	7.82	-16.86	4.34	11.82	48.44
Macro	290	8.10	7.01	8.41	-19.36	3.22	11.56	58.34
Relative Value	224	8.64	7.90	8.71	-25.89	4.58	11.91	57.46
Panel B: Fund monthly alpha (in percent, annualized)								
All Funds		3.91	3.24	8.68	-41.52	-0.54	7.73	66.93
Equity Hedge		2.90	2.37	8.56	-39.19	-1.47	6.79	57.17
Event-Driven		3.81	3.37	6.86	-19.38	0.26	6.96	36.40
Macro		5.56	4.60	9.49	-25.89	0.20	9.74	62.13
Relative Value		4.92	4.26	7.93	-27.88	1.18	8.27	46.42
Panel C: Fund characteristics								
AUM (mn USD)		201.43	45.91	539.70	0.02	12.17	160.18	8952.06
Age (months)		62.43	47.54	54.55	1.00	20.99	89.41	541.56
Management fee (%)		1.44	1.48	0.56	0.00	1.00	2.00	6.00
Performance fee (%)		18.54	20.00	5.35	0.00	20.00	20.00	47.05
Leverage (yes/no)		0.63	1.00					
Hurdle rate (yes/no)		0.13	0.00					
High watermark (yes/no)		0.91	1.00					

seven factors in Fung and Hsieh (2004).<sup>13</sup> The  $p$ -values are computed using heteroscedasticity and autocorrelation robust standard error estimators (Andrews, 1991; Andrews and Monahan, 1992). In the remainder, we denote the resulting peer performance ratio estimates for fund  $i$  in year-quarter  $q$  as  $\hat{\pi}_{i,q}^+$ ,  $\hat{\pi}_{i,q}^-$  and  $\hat{\pi}_{i,q}^0$ . Because the universe changes, the longitudinal time series of peer performance ratio estimates is unbalanced.

In Table 2 we present the summary statistics describing the universe composition and the funds' individual performance as averages over the 39 samples used to compute the peer performance ratios. Column 2 of Panel A reports the distribution of hedge funds across the different investment styles. Half of the funds classify their investments as Equity Hedge (51%). The remaining funds belong to the categories Macro (21%), Relative Value (16%) and Event-Driven (12%). Columns 3–9 of Panel A report the averages of location, scale, and shape statistics for the annualized net performance of the hedge funds in the sample. We find that, over all funds included in the universe and over all 39 samples, the average annual return is around 8.5% per year. The other statistics indicate a large cross-sectional variation in the average returns, as can be seen from the large values for the cross-sectional standard deviation (around 7.5%) and the range of annualized returns. The distribution tends to be positively skewed with an average minimum performance of -35.25% and an average top performance of 75.98%. The differences in average net performance across investment styles are small. However, the cross-sectional heterogeneity in performance is higher for the Equity Hedge funds (average cross-sectional standard deviation of 9.67% and average range of 103.93%) than for the Event-Driven hedge funds (average standard deviation of 7.82% and average range of 65.3%).

The distribution of net returns needs to be adjusted for the risk factor exposure before testing for equal performance. Panel B reports the distribution of the average (annualized) monthly alpha of the hedge funds obtained by OLS estimation of the linear model regressing the net returns in excess of the risk-free rate against the nine Carhart (1997)–Fung and Hsieh (2004) factors. We can see that adjusting for the risk factor exposure creates more heterogeneity in the average performance of the hedge funds belonging to the different investment styles. On average, the Macro hedge funds have the highest (annualized) alpha (5.56%), followed by the Relative Value funds (4.92%), and Event-Driven funds (3.81%). The worst performance regarding average alpha is by the Equity Hedge investment funds (2.90%). These differences in performance across investment styles support our choice of following investment practice by considering funds with the same style as the peer group.

The summary statistics for the size (in million US dollars), age (in months since the fund's inception date), leverage, and fee structure of the hedge funds in our database used for the regression analysis of the determinants of peer performance are presented in Panel C. We can see that the median fund in our universe is four years old at the beginning of the sample, has 45 million USD assets under management (AUM), charges a management fee of 1.48%, and has a performance fee of 20%. The median fund has a high watermark and uses leverage, but does not apply a hurdle rate.

### 3.2. Cross-section of peer performance

Let us now study the distribution of peer performance across hedge funds and how it relates to the individual performance measure of the fund. In Fig. 1, we present a two-panel plot, which in the left part displays the average (annualized) monthly alpha of the different funds sorted in descending order and grouped into 50 equal-sized buckets, where "Bucket1" corresponds to the best performing funds in terms of alpha. In the right part, a barplot displays the average estimated out-, equal-, and underperformance

<sup>13</sup> The data are retrieved from the data library of Kenneth French ([http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)) and David Hsieh (<http://faculty.fuqua.duke.edu/~dah7/DataLibrary>).

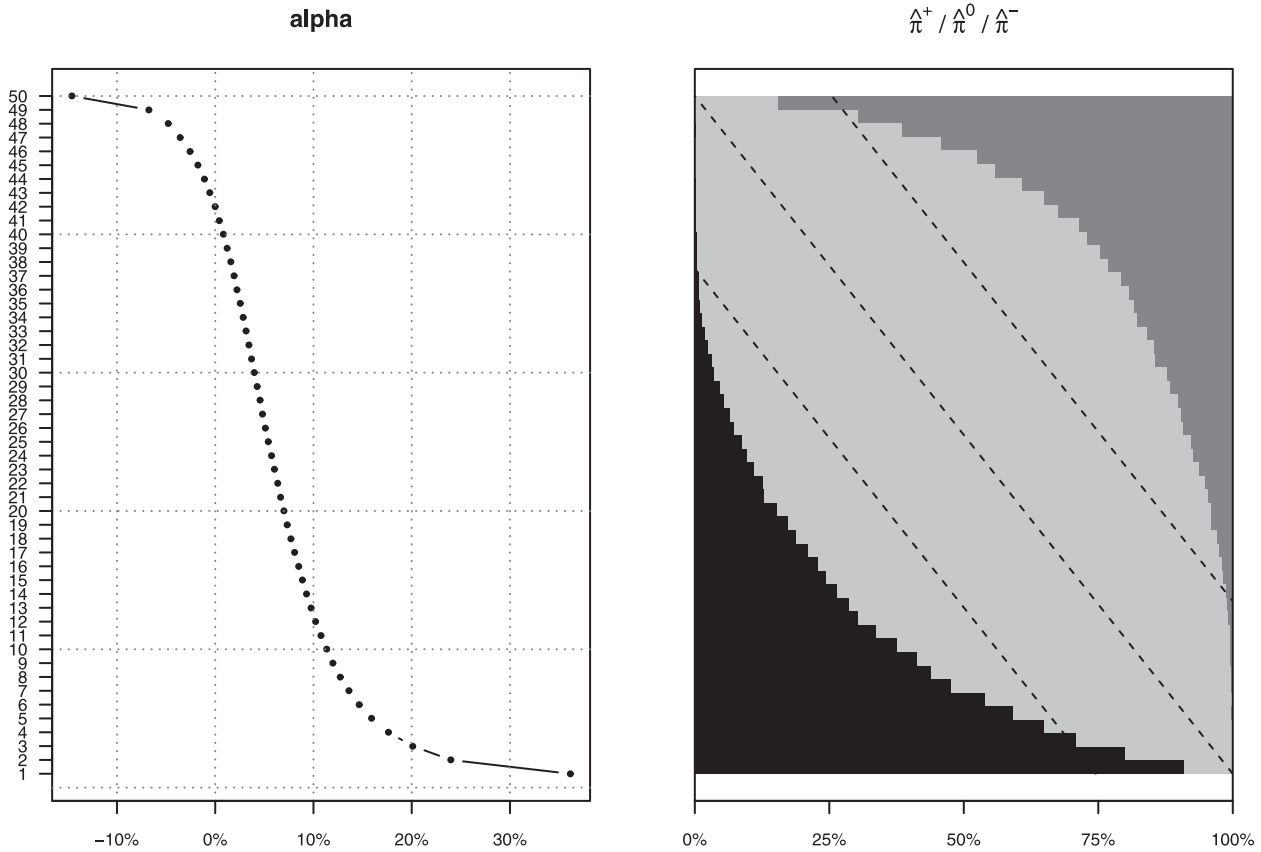


Fig. 1. Screening plot.

This figure presents the concept of screening plot. The left plot displays the average (annualized) monthly alpha for hedge funds ranked by decreasing alpha and grouped in 50 buckets. The right plot displays, for each of the 50 buckets, the average of the outperformance (black), equal-performance (light gray), and underperformance (dark gray) ratios of the hedge funds belonging to that bucket based on sorting on alpha. The diagonal dashed lines are displayed to help visualize the asymmetry in the distribution of peer performance. Average figures are computed over the 39 samples in our database (from January 2000 to June 2014). See Section 3.2 for details.

ratios in black, light gray, and dark gray, respectively. The buckets in the right plot correspond to the same 50 buckets of the left plot.

We can see that, although the alpha and the outperformance ratio are strongly positively dependent, the relationship is highly nonlinear. We find for instance that a decrease of the fund's alpha by ten percentage points has a substantially larger impact on the outperformance ratio for a top alpha performing fund than for a middle alpha performing fund. For the middle performing funds, the equal-performance ratio is above 80%, thereby indicating that they only out- or underperform a minority of their peer funds.

To screen hedge funds (as in real-life hedge fund selection), it is more usual to present the peer performance statistics in tabular form. Thus, we provide this presentation in Table 3 for funds available as of June 2014, where the performance measures are computed from July 2009 to June 2014. The funds are sorted in descending order of their alpha values. Column 1 reports the (anonymized) name of the fund. Column 2 reports the investment style. Columns 3–6 report the percentiles and the out- and under-performance ratios for the fund mentioned in the row compared to other funds with the given investment style indicated in the column header.

We notice that there is substantial scope for out- or under-performance within an investment style. For instance, four of the five underperforming funds are Equity Hedge, but the second and fifth best-performing funds also employ the Equity Hedge style. Similarly, one Macro fund is among the top-five performing hedge funds and one Macro fund is in the bottom-five performing funds.

### 3.3. Overestimation of out- and underperformance in traditional rank-based peer performance evaluation

The traditional approach to estimating the percentage of out- and underperformance is to take the percentile-rank of the fund's estimated individual performance, and to conclude that a fund with rank  $k$  outperforms  $n - k + 1$  out of  $n$  peers. This approach ignores the fact that there is also a large proportion of equal-performance between investment funds. The proposed peer performance ratios remediate this through the false discovery rate approach by explicitly taking into account the possibility of observing different estimates of the fund's individual performance, while the true performance is identical. We can thus interpret the difference between the traditional rank-based estimate of the fund's outperformance and  $\hat{\pi}_i^+$  as a proxy for the degree of overestimation in the traditional estimator:

$$\hat{\delta}_i^+ \equiv \underbrace{\text{outperformance based on ranks}}_{\equiv \left(\frac{n-k+1}{n}\right)} - \underbrace{\text{outperformance ratio}}_{\equiv \hat{\pi}_i^+} \tag{12}$$

Similarly, the overestimation of underperformance in the corresponding rank-based measure is given by:

$$\hat{\delta}_i^- \equiv \underbrace{\text{underperformance based on ranks}}_{\equiv \left(\frac{k-1}{n}\right)} - \underbrace{\text{underperformance ratio}}_{\equiv \hat{\pi}_i^-} \tag{13}$$

**Table 3**

Fund rankings and peer performance.

This table presents the fund rankings and peer performance in various peer universes for the hedge funds sample as of June 2014. It reports the five best, five most central, and five worst funds (in terms of fund's alpha), as well as the ranking and the outperformance and underperformance ratios with respect to a given universe: Equity Hedge (619 funds), Event-Driven (135 funds), Macro (254 funds) and Relative Value (201 funds). The first number is the rank of the fund within the peer universe. The square parentheses show the out- and underperformance ratios for the peer universe. The sample comprises 1,209 funds, where the performance measures are computed based on monthly figures for a five-year period ranging from July 2009 to June 2014. See Section 3.2 for details.

Fund	Strategy	Equity Hedge	Event-Driven	Macro	Relative Value
A	Relative Value	1 [0.99;0.00]	1 [1.00;0.00]	1 [1.00;0.00]	1 [1.00;0.00]
B	Equity Hedge	1 [0.98;0.00]	1 [1.00;0.00]	1 [0.97;0.00]	2 [0.94;0.00]
C	Macro	2 [0.97;0.00]	1 [1.00;0.00]	1 [0.97;0.00]	2 [0.93;0.00]
D	Relative Value	2 [0.98;0.00]	1 [1.00;0.00]	2 [0.97;0.00]	2 [0.95;0.00]
E	Equity Hedge	2 [0.86;0.00]	1 [0.70;0.00]	2 [0.69;0.00]	3 [0.35;0.00]
⋮	⋮	⋮	⋮	⋮	⋮
K	Equity Hedge	234 [0.00;0.00]	94 [0.00;0.00]	115 [0.00;0.00]	158 [0.00;0.00]
L	Equity Hedge	235 [0.21;0.00]	94 [0.00;0.26]	115 [0.17;0.05]	158 [0.00;0.38]
M	Equity Hedge	236 [0.25;0.04]	94 [0.00;0.26]	115 [0.22;0.08]	158 [0.00;0.37]
N	Equity Hedge	237 [0.00;0.00]	94 [0.00;0.00]	115 [0.00;0.00]	158 [0.00;0.08]
O	Event-Driven	238 [0.30;0.06]	94 [0.07;0.46]	115 [0.24;0.13]	158 [0.00;0.53]
⋮	⋮	⋮	⋮	⋮	⋮
V	Macro	616 [0.00;0.98]	136 [0.00;1.00]	254 [0.00;0.99]	202 [0.00;1.00]
W	Equity Hedge	616 [0.00;0.99]	136 [0.00;1.00]	255 [0.00;0.98]	202 [0.00;1.00]
X	Equity Hedge	617 [0.00;0.99]	136 [0.00;1.00]	255 [0.00;0.99]	202 [0.00;1.00]
Y	Equity Hedge	618 [0.00;0.99]	136 [0.00;1.00]	255 [0.00;1.00]	202 [0.00;1.00]
Z	Equity Hedge	619 [0.00;0.99]	136 [0.00;1.00]	255 [0.00;1.00]	202 [0.00;1.00]

Both terms can take values in the range of -100% to 100%. In practice, we expect that the percentile-rank approach overestimates both the outperformance and the underperformance, and thus that  $\hat{\delta}_i^+$  and  $\hat{\delta}_i^-$  are positive.

The left plot in Fig. 2 presents the average (over the 39 quarterly windows) estimate of the degree of overestimation in the outperformance for the 50 buckets of hedge funds in our universe, which are ranked by decreasing alpha. For all buckets considered,  $\hat{\delta}_i^+$  and  $\hat{\delta}_i^-$  are positive. On average,  $\hat{\delta}_i^+$  equals 32.28%, which indicates that, compared to the proposed outperformance ratio, the percentile-rank based approach overestimates outperformance by an average of 32.28 percentage points. Similarly, in the right plot, we can see that, compared to the underperformance ratio, the percentile-rank approach significantly overestimates underperformance by an average of 33.20 percentage points.

#### 3.4. Gains in out-of-sample portfolio performance when sorting on $\hat{\pi}^+$ and $\hat{\pi}^-$ versus $\hat{\alpha}$

Using portfolio sorts, we now analyze whether the investor benefits from selecting top quintile funds using the outperformance and underperformance ratios in comparison to using the fund's estimated alpha.<sup>14</sup> We use the top quintile portfolio in terms of the outperformance ratio to construct a portfolio of top-performing funds, and the top quintile portfolio in terms of the underperformance ratio to construct a portfolio of bottom-performing funds. We still use monthly returns for the estimation of the peer performance parameters but set the investment horizon to one quarter, six months and one year. The out-of-sample evaluation ranges from 2005-Q1 to 2014-Q2 and has thus 38 quarters. The six-month (resp. one-year) horizon portfolios are implemented as portfolios that at each quarter-end are equally invested in the current and (resp. three) previous quarter-end portfolios.

To summarize the investment value, we report in Table 4, four measures of out-of-sample portfolio performance (annualized average return, volatility, Sharpe ratio and alpha of the nine-factor model). Within the quintile portfolios, the funds are either equal-

weighted or value-weighted.<sup>15</sup> Panel A reports the result obtained when sorting using the outperformance and underperformance ratio. The performance results obtained using the alpha measures are shown in Panels B–D. In Panel B, the funds included in the top (resp. bottom) quintile portfolios are those with the highest (resp. lowest) estimated alpha. In Panel C we restrict this by the additional condition that the funds can only belong to the top and bottom quintile alpha portfolios if the corresponding  $t$ -statistic is larger than two in absolute values. In Panel D, the funds are sorted using the  $t$ -statistic of the estimated alpha.

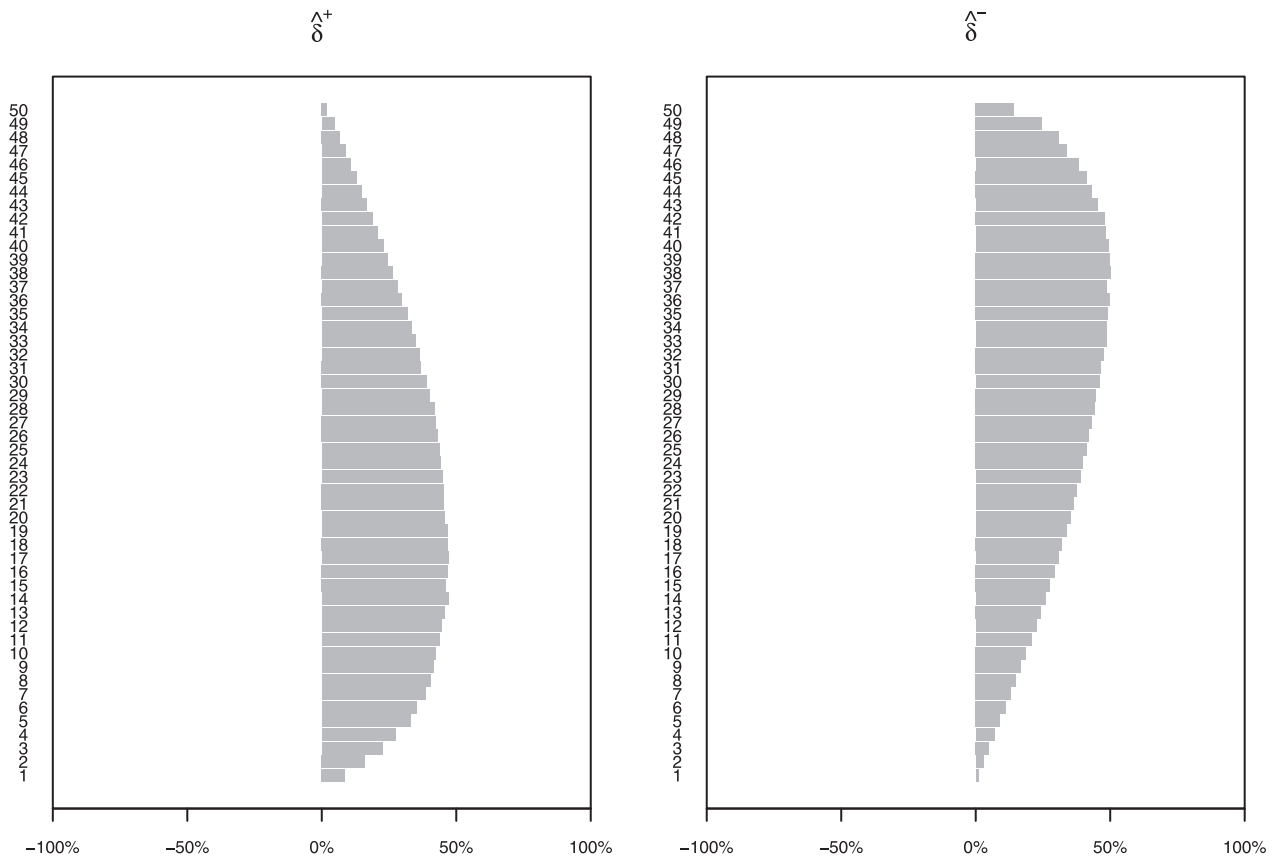
First, note that for all sorting criteria used, the difference in out-of-sample performance between the top and bottom performance portfolios is as expected. In all cases, the top quintile portfolio has a higher annualized return, lower standard deviation, higher Sharpe ratio and a higher alpha than the corresponding bottom quintile portfolio.

Second, it is of interest to note that the risk of the top quintile portfolio is substantially reduced when selecting funds based on the alpha  $t$ -statistic rather than the fund's alpha. We find that selecting funds using the alpha  $t$ -statistic leads to trade-off in terms of a lower annualized return and volatility, compared to selecting using the fund's alpha. In fact, for both the equal-weighting and value-weighting schemes and all three investment horizons, the top quintile portfolio using the alpha  $t$ -statistic has the lowest standard deviation. In contrast, the top quintile portfolio using the outperformance ratio always has the highest annualized return and alpha. This is thus *prima facie* evidence of the economic value of controlling for false discoveries when separating the good from the bad funds in a universe of peer funds.

As a robustness test, we present in Panel E of Table 4, the results obtained when the peer group is enlarged to all funds in the universe, instead of restricting it to the funds belonging to the same investment style. Statistically, this can improve the accuracy of the estimation, but economically, it creates a possible "apples and oranges" problem since we mix funds with different investment styles in the peer analysis. Comparing Panel A

<sup>14</sup> As a robustness check, we verified that the same conclusions were obtained using decile portfolios instead of quintile portfolios.

<sup>15</sup> In case of value-weighting, the lagged value of the fund's assets under management is used. The advantage of considering value-weighting compared with equal-weighting is that the performance is not driven by small hedge funds that may not be investable.



**Fig. 2.** The degree of overestimation in rank-based estimation of outperformance and underperformance.

The outperformance (underperformance) ratio is a robust alternative to the common practice of measuring outperformance (underperformance) as the percentage of peer funds for which the fund's performance is lower (higher) than the fund of interest. We interpret the difference between the rank-based outperformance (underperformance) and the outperformance (underperformance) ratio as an estimate for the degree of overestimation in the rank-based outperformance. This figure presents the average values of the overestimation for the hedge funds in our database ranked by decreasing alpha and grouped in 50 buckets. The left (right) plot displays the overestimation in percentile-rank estimates of the outperformance (underperformance) ratio. Peer performance ratios used to compute  $\hat{\delta}^+$  and  $\hat{\delta}^-$  are based on the universe of funds pursuing the same strategy. Average figures are computed based on the 39 samples in our database (from January 2000 to June 2014). See Section 3.3 for details.

with Panel E, we effectively find suggestive evidence that a better out-of-sample performance is obtained by restricting the funds to the same investment style. This confirms our recommendation to control for fund style in the peer analysis and is consistent with Hunter et al. (2014). A detailed investigation of the optimal choice of a peer group is an interesting avenue for further research.<sup>16</sup>

Overall, we find that the outperformance ratio is effectively able to select the top performing hedge funds. In the next section, we analyze this outcome in further detail and use multivariate regression techniques to compare the complementarity with alternative peer performance measures and to control for other influences in the analysis of the predictive power of the outperformance ratio for forecasting fund performance over the next one-, two- and four quarters.

<sup>16</sup> As an alternative to defining peers based on the reported investment style, one could use a data-driven procedure to identify peer funds. Broadly speaking, we can distinguish two classes of data-driven procedures to identify peer funds. One approach uses historical returns and fund characteristics (e.g., age, size, fund style and obtains the peer funds through cluster analysis and fund matching algorithms, or via an analysis of the coefficients in a regression of fund returns on benchmark indices (see, e.g., Brown and Goetzmann, 1997). The second approach classifies funds based on the highest percentage of overlap in fund holdings (Cremers and Petajisto, 2009). In practice, investors usually define the peer funds as the set of other funds in a portfolio, funds in a screening list of analysts (e.g., those followed by the investment committee), or funds with the same investment style.

### 3.5. Complementarity of $\hat{\pi}^+$ to alternative peer performance measures

Various alternatives exist to create quintile portfolios in terms of outperforming funds. We examine now the added value of combining the information in the outperformance ratio-based top quintile portfolio with quintile portfolios constructed using other performance measures. We verify the complementarity in terms of the difference in composition, as well as in a regression analysis quantifying the expected additional return the hedge fund yields when belonging to the top quintile portfolio, compared with the reference hedge fund that is not included in the top quintile portfolio.

We consider four alternative fund performance-based selection measures: (i) the fund's *past return* capturing the "hot hands effect" (see Agarwal and Naik, 2000), (ii) the fund's *alpha*, (iii) the fund's *relative alpha*, and (iv) the fund's *peer alpha*. The fund's *relative alpha* ( $\hat{\alpha}^{rel}$ ), defined by Jagannathan et al. (2010), is computed as the OLS estimate of the intercept obtained in the regression of the hedge fund's returns on the US aggregate market factor and the self-reported style factor as explanatory variables together with an additional variable selected by Bayesian information criterion (BIC) among the (lagged) values of HFRI indices.<sup>17</sup> Finally, the *peer alpha*

<sup>17</sup> Overall, we use the 37 HFRI style factors corresponding to the HFR styles described in <https://www.hedgefundresearch.com/hfr-hedge-fund-strategy-classification-system>.

**Table 4**

Out-of-sample performance results of the quintile portfolios.

This table presents the annualized return (Mean, in percent), volatility (Std, in percent), Sharpe ratio and nine-factor model alpha (in percent) of the quarterly rebalanced portfolios invested in the top and bottom quintile of the hedge funds sorted by various peer performance measures. When the investment horizon  $H$  is semi-annual (S) or annual (A), funds entering the portfolio at year-quarter  $q$  can only be disinvested in year-quarter  $q + 2$  and  $q + 4$ , respectively. Panel A reports results for portfolios invested in top quintiles of the hedge funds ranked on their outperformance ( $\hat{\pi}^+$ ) and underperformance ratios ( $\hat{\pi}^-$ ) computed using funds with the same strategy as peers. Panels B and C report results for portfolios invested in the bottom and top quintile of the hedge funds ranked on their alpha ( $\hat{\alpha}$ ) and ranked on their alpha for which the  $t$ -statistic is higher than two in absolute values, respectively. Panel D reports results for portfolios invested in the bottom and top quintile of the hedge funds ranked on their alpha  $t$ -statistic. Panel E reports results for portfolios invested in the quintile of the hedge funds ranked on their outperformance and underperformance ratios computed using all funds in the universe as peers. Peer performance measures are computed on five-year rolling samples of monthly net returns. Column "Alpha" reports the annualized quarterly alpha (in percent) of the portfolios excess return against the nine-factor model. The symbols \*\*\*, \*\* and \* indicate statistical significance of the strategy's alpha at the 1%, 5%, and 10% levels, respectively, based on heteroscedasticity and autocorrelation robust standard error estimators. The out-of-sample evaluation ranges from 2005–Q1 to 2014–Q2 for a total of 38 rebalancing dates. See Section 3.4 for details.

H	Quintile	Equally-weighted portfolios				Value-weighted portfolios			
		Mean	Std	Sharpe	Alpha	Mean	Std	Sharpe	Alpha
Panel A: Portfolio sorts using the outperformance and underperformance ratio									
Q	Top $\hat{\pi}^+$	9.39	8.14	1.15	5.52***	7.54	8.17	0.92	3.72**
Q	Top $\hat{\pi}^-$	4.74	10.59	0.45	0.01	3.81	9.05	0.42	0.24
S	Top $\hat{\pi}^+$	8.80	7.86	1.12	4.92***	7.08	7.88	0.90	3.38**
S	Top $\hat{\pi}^-$	5.53	9.99	0.55	0.90	4.73	8.59	0.55	1.36
A	Top $\hat{\pi}^+$	7.99	7.95	1.01	4.16**	6.56	8.03	0.82	2.98*
A	Top $\hat{\pi}^-$	5.77	8.82	0.65	1.15	5.02	7.66	0.66	1.57
Panel B: Portfolio sorts using alpha									
Q	Top	8.98	8.82	1.02	4.78**	6.83	8.47	0.81	2.64
Q	Bottom	5.80	11.72	0.49	0.13	5.66	10.77	0.53	0.86
S	Top	8.68	8.53	1.02	4.47**	6.63	8.26	0.80	2.60
S	Bottom	6.54	11.08	0.59	0.93	6.77	9.89	0.68	2.27
A	Top	7.88	8.45	0.93	3.72**	6.12	8.40	0.73	2.34
A	Bottom	6.75	9.65	0.70	1.21	6.79	8.71	0.78	2.18*
Panel C: Portfolio sorts using significant alpha ( $ t\text{-stat}  > 2$ )									
Q	Top	7.01	6.15	1.14	4.27***	6.20	6.17	1.00	3.37**
Q	Bottom	5.46	8.95	0.61	0.93	5.44	8.16	0.67	1.59
S	Top	6.89	5.84	1.18	4.20***	6.00	6.07	0.99	3.36**
S	Bottom	5.81	8.82	0.66	1.33	5.48	8.16	0.67	1.56
A	Top	6.31	5.78	1.09	3.58***	5.48	6.13	0.89	2.76**
A	Bottom	5.75	8.36	0.69	1.25	5.41	7.89	0.69	1.44
Panel D: Portfolio sorts using alpha $t$ -stat									
Q	Top	7.17	5.62	1.27	4.55***	6.05	6.24	0.97	3.22**
Q	Bottom	5.70	10.83	0.53	0.25	5.74	9.22	0.62	1.23
S	Top	6.83	5.38	1.27	4.27***	5.81	5.98	0.97	3.16**
S	Bottom	6.16	10.41	0.59	0.70	6.21	8.96	0.69	1.74
A	Top	6.18	5.56	1.11	3.58***	5.37	6.16	0.87	2.71**
A	Bottom	6.39	9.25	0.69	1.04	6.13	7.95	0.77	1.77*
Panel E: Portfolio sorts using peer ratios computed with all funds as peers									
Q	Top $\hat{\pi}^+$	8.87	7.86	1.13	5.00***	6.76	7.77	0.87	2.97*
Q	Top $\hat{\pi}^-$	5.5	10.74	0.51	0.37	5.27	9.28	0.57	1.16
S	Top $\hat{\pi}^+$	8.30	7.60	1.09	4.48***	6.41	7.54	0.85	2.78*
S	Top $\hat{\pi}^-$	6.13	10.29	0.60	1.03	6.14	8.88	0.69	2.23*
A	Top $\hat{\pi}^+$	7.51	7.65	0.98	3.81**	5.92	7.72	0.77	2.48
A	Top $\hat{\pi}^-$	6.37	9.10	0.70	1.36	6.31	7.86	0.80	2.38**

( $\hat{\alpha}^{\text{peer}}$ ), as suggested by Hunter et al. (2014), is the OLS estimate of the intercept of the same regression as used to obtain the fund's alpha, except that the factor corresponding to the average return of the funds in the same investment style is included as an active peer benchmark factor.<sup>18</sup>

A necessary condition for complementarity of the top quintile portfolios is that there is no perfect overlap in their composition. We examine this for the outperformance ratio by reporting the

percentage of funds that belong to the top quintile portfolio constructed using the alternative measures, when they also belong to the top quintile portfolio for the outperformance ratio. We compute this percentage for each quarter in our sample. The overlap is the highest for the fund's alpha measure. It varies between 75% and 90% with an average value of 82%. The lowest overlap is with respect to the fund's lagged return, for which the overlap ranges between 15% and 40% with an average value of 30%. For the relative and peer alpha, we find that the overlap with the outperformance ratio-based top quintile portfolio is similar. In both cases, it is on average 70%, and the range is between 55% and 80%.

We use regression methods to quantify the stand-alone predictive value of the individual portfolio sorts when forecasting the fund's (risk-adjusted) return, as well as the incremental predictive value of the portfolio sorts using the outperformance ratio, as compared to the use of alternative peer performance measures. The variable to forecast, denoted by  $Y_{i,q}$ , is either the fund  $i$ 's future return or the future risk-adjusted excess return, computed as the

<sup>18</sup> We also tested alternatives using the *distinctiveness* and the *selectivity* peer performance measures proposed by Sun et al. (2012) and Amihud and Goyenko (2013), respectively. These measures associate outperformance to funds for which the track record or returns cannot be explained by their peer returns or the factor model. They do not take the direction of the performance differential into account, which, under the model of Titman and Tiu (2011) can be optimal when funds only take active positions when they have positive skills. Several empirical tests on our sample of hedge funds indicate, however, that in practice these measures have no significantly positive association with future fund performance. To save space, we therefore, focus on the peer performance measures that do take the direction of the relative performance into account.

fund's return in excess of the risk-free rate, and expressed in units per risk by dividing it by the standard deviation of the preceding 12 monthly returns. The future returns are computed over a quarterly, semi-annual and annual horizon. Because of the high overlap between the top quintile alpha portfolio funds and the top quintile outperformance ratio funds, and to avoid identification issues due to quasi-multicollinearity, we do not consider the complementarity with respect to the fund's alpha in the regression analysis.

The models we estimate to predict, at year-quarter  $q$ , the future performance  $Y_{i,q}$ , are nested in the following specification:

$$Y_{i,q} = \zeta_q + \beta_1 TQ_{i,q}^{\hat{\pi}^+} + \beta_2 TQ_{i,q}^R + \beta_3 TQ_{i,q}^{\hat{\alpha}^{\text{rel}}} + \beta_4 TQ_{i,q}^{\hat{\alpha}^{\text{peer}}} + \gamma' \mathbf{CTRL}_{i,q} + \varepsilon_{i,q}, \quad (14)$$

where  $TQ_{i,q}^{\hat{\pi}^+}$  is the dummy variable indicating that fund  $i$  belongs to the 20% funds with highest outperformance ratio in year-quarter  $q$ . The variables  $TQ_{i,q}^R$ ,  $TQ_{i,q}^{\hat{\alpha}^{\text{rel}}}$  and  $TQ_{i,q}^{\hat{\alpha}^{\text{peer}}}$  are defined similarly, but using the alternative performance measure.<sup>19</sup> We use  $\varepsilon_{i,q}$  to denote the error term, and  $\mathbf{CTRL}_{i,q}$  to refer to the inclusion of control variables. As control variables, we include the assets under management (in logarithm), the fund's age (in months, since inception), and the fund's capital inflow (in thousand USD) defined by [Fung et al. \(2008\)](#). We further control for risk in the return prediction model by including the fund's return volatility (computed as the standard deviation over the last twelve months) and the time-invariant fund characteristics:

$$\mathbf{CTRL}_{i,q} \equiv (LAUM_{i,q}, AGE_{i,q}, F_{i,q}, VOL_{i,q}, EH_i, MA_i, RV_i, MF_i, PF_i, LEV_i, HUR_i, HWM_i)',$$

where  $EH_i$ ,  $MA_i$  and  $RV_i$  are dummies indicating the Equity Hedge, Macro, and Relative Value hedge fund styles (Event-Driven is the reference category).<sup>20</sup>  $MF_i$  and  $PF_i$  are the fund's management and performance fees. The indicator variable  $LEV_i$  is one if the fund is allowed to use leverage.  $HUR_i$  is another dummy variable, which indicates the presence of a required rate of return that the fund manager needs to obtain before collecting the performance fee. The dummy variable  $HWM_i$  indicates that the fund has a high watermark provision that requires the manager to make up past deficits before earning the incentive fee. All of these variables are standard in previous studies of fund performance (see, e.g., [Liang, 1999](#)).

Let us first discuss the stand-alone predictive value of being a top quintile performing fund, as estimated by the univariate regression that forecasts  $Y_{i,q}$  using only one dummy variable  $TQ_{i,q}^*$ . The corresponding coefficient  $\beta$ , expresses then the predicted gain in (risk-adjusted excess) return when the fund belongs to the top quintile portfolio, compared to the funds that do not belong to the top quintile portfolio. The OLS estimation results are reported in the left (resp. right) part of [Table 5](#), when the variable of interest is the quarterly (Panel A), semi-annual (Panel B) and annual (Panel C) fund return (resp. risk-adjusted excess return). Based on the magnitude of the coefficient and the adjusted  $R^2$  of the predictive regressions, we find that the top quintile dummy using the outperformance ratio is always the second best. For forecasting the return, the top momentum quintile is best, while for forecasting the risk-adjusted return, the top peer alpha quintile yields the highest increase in risk-adjusted return. The top quintile based on the outperformance ratio thus strikes a balance between the objectives of

achieving both a high expected return and a high risk-adjusted expected return. We further find that the expected increases in (risk-adjusted) expected return is higher for longer horizons, which is consistent with the time-scaling of expected returns over longer horizons.

In practice, we recommend to use the outperformance ratio in combination with other predictors of (risk-adjusted) fund performance. This advice follows from our finding that the outperformance ratio has incremental predictive value when included in multivariate regressions. Consider first the multivariate regressions results in [Table 5](#) with two quintile portfolio dummies and no other control variables. Suppose the corresponding coefficients are  $\beta_1$  and  $\beta_2$ . In case of complementarity, both  $\beta_1$  and  $\beta_2$  need to be positive such that the predicted gain in (risk-adjusted) excess return (namely  $\beta_1 + \beta_2$ ) is higher than the expected gain when the fund belongs to only one of the top quintile portfolios. This hypothesis is confirmed by estimates reported in [Table 5](#). In all models considered, the estimates of  $\beta_1$  and  $\beta_2$  are positive and, except for the annual investment horizon, statistically significant. This thus indicates that the fund's outperformance ratio, the fund's lagged return, and the fund's relative and peer alphas are good predictors of future fund performance, and that the predictive power of one is not cannibalized by the others.

The previous analysis indicates that the ability of hedge funds to outperform their peers is too complex to be captured by only a single performance measure. In addition to the use of multiple top quintile dummies, other variables such as the fund size and risk are relevant. For brevity, we focus the presentation of our results to the forecasting at the quarterly prediction horizon.<sup>21</sup> Results are reported in [Table 6](#). Note that fund size and the fund's performance fee have significant negative effects on future firm (net) return performance. For our sample, we find that the higher the fund risk, the higher the predicted return, but the lower is the expected risk-adjusted return. These results are in line with previous research. Note also that all the coefficients on the top quintile portfolio dummies are positive and significant at the 1% significance level. The main take-away finding of [Table 6](#) is thus that including those extra control variables does not remove the statistical significance of the finding that a fund belonging to the top quintile portfolio in terms of the outperformance ratio can be expected to have a higher (risk-adjusted) return. Further, this effect is not explained by the usual predictors of fund performance.

### 3.6. Determinants of peer performance

In this section, we conduct additional tests to verify well-established theories about the effects of fund size and fund age on the peer performance of hedge funds. Regarding fund size, we expect to find decreasing returns to scale in active management, and thus that fund size deteriorates the peer performance of hedge funds, that is, it decreases the outperformance ratio and increases the underperformance ratio. [Chen et al. \(2004\)](#) attribute the dis-economy of scale to the fact that small hedge funds have more opportunities to deploy their talent, especially when investing in less liquid assets.

A second channel through which fund size impacts on peer performance is reputation risk, which is highest for funds with both a long track record and a large amount of assets under management. In fact, in addition to performance-based compensation, hedge funds receive compensation in the form of management fees, which are directly related to fund size. If there

<sup>19</sup> All dummies are computed using quarterly performance measures, except for  $TQ_{i,q-1}^R$ , which is the dummy for the funds with the highest return over the past 1, 2 and 4 quarters when  $Y_{i,q}$  is the (risk-adjusted) return over the quarterly, semi-annual and annual horizon, respectively.

<sup>20</sup> All hedge funds considered are US funds, so we do not include a dummy variable for the fund domicile (see [Aragon et al., 2013](#)).

<sup>21</sup> Similar results are obtained at the semi-annual forecasting horizon. Like for the regression without the control variables, the statistical evidence for the complementarity of the outperformance ratio and the alternative measure disappears at the annual forecasting horizon.

**Table 5**

Return prediction results.

This table presents the regression results of the prediction model (14) without control variables when the dependent variable is either the future net return (left part) or the future risk-adjusted excess return (right part). The model aims at quantifying the stand-alone predictive value (i.e., univariate regressions) and the incremental predictive value (i.e., multivariate regressions) that a fund is in the top quintile portfolio according to the fund's outperformance ratio ( $TQ^{\pi^+}$ ), the fund's lagged return ( $TQ^R$ ), the fund's relative alpha ( $TQ^{\alpha^{rel}}$ ), and the fund's peer alpha ( $TQ^{\alpha^{peer}}$ ). The respective regression coefficients are to be interpreted as expected gains in performance when the fund belongs to the top quintile portfolio, compared to the reference group of hedge funds that do not belong to either of the top quintile groups included in the regression. Panel A reports results for the quarterly prediction horizon, Panel B for the semi-annual horizon and Panel C for the annual horizon. Robust standard errors are shown in parentheses. The symbols \*\*\*, \*\* and \* indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Adj  $R^2$  reports the (adjusted) R-squared (in percent). See Section 3.5 for details.

	Return								Risk-adjusted excess return									
	Univariate regressions				Multivariate regressions				Univariate regressions				Multivariate regressions					
Panel A: Quarterly prediction horizon																		
$TQ^{\pi^+}$	0.83***				0.62***				0.75***	0.71***	0.25***					0.24***	0.15***	0.14***
	(0.11)				(0.10)				(0.14)	(0.14)	(0.03)					(0.03)	(0.03)	(0.04)
$TQ^R$		1.90***				1.83***	1.86***	1.86***				0.10***				0.08***	0.08***	0.08***
		(0.11)				(0.11)	(0.11)	(0.11)				(0.02)				(0.02)	(0.02)	(0.02)
$TQ^{\alpha^{rel}}$			0.59***				0.40***		0.13			0.25***				0.24***		0.16***
			(0.11)				(0.11)		(0.14)			(0.03)				(0.03)		(0.04)
$TQ^{\alpha^{peer}}$				0.64***				0.45***		0.20			0.26***			0.25***		0.18***
				(0.11)				(0.11)		(0.14)			(0.03)			(0.03)		(0.04)
Adj $R^2$	0.16	0.84	0.08	0.09	0.93	0.88	0.89	0.16	0.16	0.51	0.09	0.51	0.56	0.55	0.56	0.61	0.63	0.66
Panel B: Semi-annual prediction horizon																		
$TQ^{\pi^+}$	1.18***				0.81***				1.03***	1.00***	0.29***					0.27***	0.19***	0.18***
	(0.22)				(0.22)				(0.26)	(0.26)	(0.04)					(0.04)	(0.05)	(0.05)
$TQ^R$		2.49***				2.36***	2.41***	2.41***				0.19***				0.14***	0.15***	0.15***
		(0.20)				(0.20)	(0.19)	(0.19)				(0.03)				(0.03)	(0.03)	(0.03)
$TQ^{\alpha^{rel}}$			0.87***				0.54**		0.25			0.29***				0.27***		0.17***
			(0.23)				(0.23)		(0.27)			(0.04)				(0.04)		(0.05)
$TQ^{\alpha^{peer}}$				0.90***				0.59***		0.28			0.29***			0.27***		0.18***
				(0.23)				(0.22)		(0.27)			(0.04)			(0.04)		(0.05)
Adj $R^2$	0.12	0.56	0.07	0.07	0.62	0.59	0.59	0.13	0.13	0.53	0.21	0.51	0.52	0.65	0.64	0.65	0.64	0.64
Panel C: Annual prediction horizon																		
$TQ^{\pi^+}$	1.00**				0.49				0.72	0.74	0.26***					0.23***	0.16**	0.16**
	(0.44)				(0.44)				(0.51)	(0.51)	(0.06)					(0.06)	(0.06)	(0.06)
$TQ^R$		2.45***				2.34***	2.36***	2.37***				0.17***				0.12***	0.13***	0.13***
		(0.40)				(0.41)	(0.40)	(0.40)				(0.04)				(0.04)	(0.04)	(0.04)
$TQ^{\alpha^{rel}}$			0.91**				0.45		0.47			0.26***				0.23***		0.16**
			(0.45)				(0.45)		(0.52)			(0.06)				(0.06)		(0.07)
$TQ^{\alpha^{peer}}$				0.88*				0.42		0.42			0.25***			0.23***		0.16**
				(0.45)				(0.45)		(0.53)			(0.06)			(0.06)		(0.07)
Adj $R^2$	0.03	0.22	0.03	0.03	0.23	0.22	0.22	0.04	0.04	0.27	0.12	0.27	0.27	0.32	0.33	0.33	0.33	0.33

are more assets under management, the hedge fund risks losing more due to a loss of reputation when failing unconventionally. The herding models proposed by Scharfstein and Stein (1990) and Graham (1999) predict that managers with high reputation and salary (which are associated with higher assets under management) tend to herd more. This career hypothesis has been recently confirmed by Boyson (2010) based on hedge fund data.

We use a nonlinear regression framework to test these hypotheses in our longitudinal time series of quarterly updated peer performance ratios  $\hat{\pi}_{i,q}^+$ ,  $\hat{\pi}_{i,q}^-$  and  $\hat{\pi}_{i,q}^0$  ( $q = 1, \dots, 39$ ). Under this approach, the expectation of the peer performance ratio is modeled as a logistic function of the fund's assets under management at the end of the preceding quarter (in logarithm,  $LAUM_{i,q-1}$ ), the age of the fund at the end of the preceding quarter (i.e., time since inception in months,  $AGE_{i,q-1}$ ), an interaction term (i.e.,  $IT_{i,q-1} \equiv LAUM_{i,q-1} \times AGE_{i,q-1}$ ), and control variables:

$$\hat{\pi}_{i,q}^* = G(\beta_{0,q} + \beta_{1,q}LAUM_{i,q-1} + \beta_{2,q}AGE_{i,q-1} + \beta_{3,q}IT_{i,q-1} + \gamma_q'CTRL_i) + \varepsilon_{i,q}, \tag{15}$$

where  $G(\cdot)$  is the logistic function,  $\varepsilon_{i,q}$  is the error term,  $\gamma_q$  is a  $(8 \times 1)$  vector of parameters for the control variables  $CTRL_i \equiv (EH_i, MA_i, RV_i, MF_i, PF_i, LEV_i, HUR_i, HWM_i)'$ , as defined in Section 3.5.

The dependent variable  $\hat{\pi}_{i,q}^*$  is either the outperformance ratio  $\hat{\pi}_{i,q}^+$ , the underperformance ratio  $\hat{\pi}_{i,q}^-$ , or the equal-performance ratio  $\hat{\pi}_{i,q}^0$  of fund  $i$  at quarter  $q$ .<sup>22</sup> Model (15) is estimated by nonlinear least squares for each of the 39 estimation samples separately.<sup>23</sup> The regression results are presented in Panel A of Table 7 for the separate models explaining the outperformance ratio (Columns 2–5), the underperformance ratio (Columns 6–9), and the equal-performance ratio (Columns 10–13). For each dependent variable, we report the average value (over the 39 samples) of the nonlinear regression coefficients and the percentage of the samples for which the estimated coefficient is significantly different

<sup>22</sup> A potentially more efficient alternative to the equation-by-equation estimation for  $\hat{\pi}_i^+$ ,  $\hat{\pi}_i^-$  and  $\hat{\pi}_i^0$  is to estimate them jointly by maximum likelihood under a Dirichlet distribution that considers the fact that the three ratios add up to unity. Another alternative is to specify a multinomial logit approach for the out-, equal-, or underperformance status of each pair of funds. However, these multivariate models pose considerable econometric challenges regarding the model for the dependence between the peer performance ratios.

<sup>23</sup> We also considered estimating the model by panel estimation methods by assuming that the slope coefficients are identical across the 39 samples and including fund fixed effects. However, because of the nonlinear regression specification, this requires the estimation of the fixed effects and slope coefficients jointly, which leads to the curse of dimensionality during the estimation process. The alternative to linearizing the regression by using  $G^{-1}(\hat{\pi}_{i,q}^*)$  is also not possible because of the non-negligible proportion of funds for which  $\hat{\pi}_{i,q}^*$  is exactly one.

**Table 6**

Return prediction results with control variables.

This table presents the regression results of the return prediction model (14) for the quarterly prediction horizon. The dependent variable is either the quarterly net return (left part) or the quarterly risk-adjusted excess return (right part) of hedge fund  $i$  at the end of quarter  $q$ . The model aims at quantifying the complementarity of the information value that a fund is in the top quintile portfolio according to the fund's outperformance ratio ( $TQ^{\bar{\alpha}^+}$ ), the fund's lagged return ( $TQ^R$ ), the fund's relative alpha ( $TQ^{\bar{\alpha}^{rel}}$ ), and (iii) the fund's peer alpha ( $TQ^{\bar{\alpha}^{peer}}$ ). The control variables are the fund's asset under management (in logarithm,  $LAUM$ ), the fund's age  $AGE$ , the fund's capital inflow  $F$  (in thousand USD), the fund's return volatility (computed as the standard deviation over the last twelve months,  $VOL$ ). Variables  $EH$ ,  $MA$ , and  $RV$  are dummies indicating the Equity Hedge, Macro, and Relative Value hedge fund styles, respectively.  $MF$  and  $PF$  are the fund's management and performance fees.  $LEV$ ,  $HUR$ , and  $HWM$  are dummy variables indicating that the fund uses leverage, hurdle rate, and high watermark provision, respectively. Robust standard errors are shown in parentheses. The symbols \*\*\*, \*\* and \* indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Adj  $R^2$  reports the adjusted R-squared. See Section 3.5 for details.

	Return				Risk-adjusted excess return			
$TQ^{\bar{\alpha}^+}$	0.57*** (0.10)	0.46*** (0.10)	0.36*** (0.12)	0.30** (0.12)	0.24*** (0.03)	0.22*** (0.03)	0.13*** (0.03)	0.11*** (0.03)
$TQ^R$		1.16*** (0.10)				0.19*** (0.02)		
$TQ^{\bar{\alpha}^{rel}}$			0.35*** (0.12)				0.18*** (0.03)	
$TQ^{\bar{\alpha}^{peer}}$				0.44*** (0.12)				0.20*** (0.03)
$LAUM$	-0.06** (0.03)	-0.06** (0.03)	-0.06** (0.03)	-0.06** (0.03)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
$AGE$	-0.02* (0.01)	-0.01* (0.01)	-0.01 (0.01)	-0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
$F$	-0.25 (0.22)	-0.21 (0.21)	-0.25 (0.22)	-0.25 (0.22)	-0.06 (0.04)	-0.06 (0.04)	-0.07 (0.04)	-0.07 (0.04)
$VOL$	0.44*** (0.02)	0.40*** (0.02)	0.44*** (0.02)	0.45*** (0.02)	-0.04*** (0.01)	-0.04*** (0.00)	-0.04*** (0.00)	-0.04*** (0.00)
$EH$	0.47*** (0.12)	0.50*** (0.12)	0.49*** (0.12)	0.49*** (0.12)	0.22*** (0.05)	0.23*** (0.05)	0.23*** (0.05)	0.23*** (0.05)
$MA$	-0.88*** (0.11)	-0.84*** (0.10)	-0.88*** (0.11)	-0.89*** (0.11)	-0.22*** (0.04)	-0.22*** (0.04)	-0.23*** (0.04)	-0.23*** (0.04)
$RV$	0.53*** (0.11)	0.56*** (0.11)	0.54*** (0.11)	0.54*** (0.11)	0.40*** (0.04)	0.40*** (0.04)	0.40*** (0.04)	0.40*** (0.04)
$MF$	-5.63 (7.12)	-6.12 (6.90)	-5.96 (7.09)	-6.00 (7.08)	-0.17 (3.07)	-0.25 (3.06)	-0.34 (3.06)	-0.34 (3.05)
$PF$	-1.53* (0.87)	-1.43* (0.85)	-1.68* (0.87)	-1.72** (0.87)	-0.13 (0.36)	-0.11 (0.36)	-0.21 (0.36)	-0.21 (0.36)
$LEV$	-0.11 (0.08)	-0.11 (0.08)	-0.11 (0.08)	-0.11 (0.08)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)
$HUR$	0.11 (0.12)	0.10 (0.11)	0.11 (0.12)	0.11 (0.12)	0.09* (0.05)	0.08* (0.05)	0.09* (0.05)	0.09* (0.05)
$HWM$	0.08 (0.16)	0.07 (0.15)	0.08 (0.16)	0.07 (0.16)	0.09 (0.07)	0.09 (0.07)	0.09 (0.07)	0.09 (0.07)
Adj $R^2$	20.58	20.86	20.59	20.60	16.23	16.51	16.40	16.43

from zero, significantly positive, and significantly negative at the 5% level, respectively.

Let us first discuss the effects of fund size, fund age, and the interaction effect between fund size and fund age. For the outperformance ratio, we find that for 77% of the samples, the coefficient of fund size is significantly negative. The interaction effect between fund size and fund age is not significantly different from zero in 87% of the samples. It follows that consistent with the presence of decreasing returns to scale, fund size leads to a deterioration in the fund's outperformance ratio. In 33% of the samples, fund age is associated with a lower outperformance ratio.

For the underperformance ratio, the effects of fund size and fund age are more complex because the interaction variable between fund age and fund size is significantly negative in 67% of the samples. We find that the coefficient on fund size and fund age is significantly positive for 62% and 97% of the samples, respectively, while the coefficient on the interaction variable between fund size and fund age is significantly negative for 67% of the samples. It follows that we have the following three *ceteris paribus* interpretations. First, for two funds of the same size, the oldest fund tends to have a higher underperformance ratio. This result confirms the analysis of Aggarwal and Jorion (2010) who find that hedge funds tend to add value in their early years and that thereafter the per-

formance tends to deteriorate in a nonlinear manner. Second, we have that for two funds of the same age, the largest fund tends to have the highest underperformance ratio. Third, we find that the increase in the underperformance ratio for larger (resp. older) funds is partly compensated for by the age (resp. size) of the fund. In other words, large funds with a long track tend to underperform less than their younger peers with same fund size.

Finally, for the equal-performance ratio, the interaction is almost never significant, whereas fund size has a positive impact on the equal-performance ratio. These two results are consistent with the prediction that, because of decreasing returns to scale in active investment strategies, larger funds have fewer opportunities to outperform their smaller peers.

Regarding the fund characteristics, we find that funds with an Equity Hedge and Macro investment style tend to have a higher equal-performance ratio and a lower out- and underperformance ratio than the Event-Driven funds. More interestingly, for 46% and 87% of the samples, the value of the performance and management fee is associated with a higher outperformance ratio. Similarly, for most of the samples, the level of the fee decreases the underperformance ratio. Since we compute these ratios on net returns, it follows that for our universe of hedge funds, higher fees tends to be associated with better performing funds, which is consistent

**Table 7**

Nonlinear regression analysis of the cross-section of hedge funds' peer performance ratios.

This table presents the summary of the estimation results for the nonlinear model (15) with the dependent variable  $\hat{\pi}^*$  and  $\hat{\delta}^*$  (computed over the 39 samples). Panel A reports results for peer ratios computed using funds with the same strategy as peers (i.e.,  $\hat{\pi}^+$ ,  $\hat{\pi}^-$  and  $\hat{\pi}^0$ ). Panel B reports results for the determinants of the overestimation in outperformance and underperformance when using the percentile-rank approach (i.e.,  $\hat{\delta}^+$  and  $\hat{\delta}^-$ ). For each peer performance ratio, the table shows the average value  $\bar{\beta}$  (over the 39 samples) of the regression coefficients and the percentage of samples for which the estimated coefficient is significantly different from zero ( $\neq 0$ ), significantly positive ( $> 0$ ), and significantly negative ( $< 0$ ) at the 5% level, respectively (where we use asymptotic standard errors for the tests). The main variables of interest are the fund's assets under management (in logarithm, *LAUM*), the fund's age in months (*AGE*), and their interaction ( $IT \equiv LAUM \times AGE$ ). The other control variables are described in Table 6 (page 38). Pseudo  $R^2$  (in percent) reports the average of the ratio between the standard deviation of the errors and the standard deviation of the dependent variable. See Section 3.6 for details.

Panel A: Determinants of peer performance ratios												
	$\hat{\pi}^+$				$\hat{\pi}^-$				$\hat{\pi}^0$			
	$\bar{\beta}$	$\neq 0$	$> 0$	$< 0$	$\bar{\beta}$	$\neq 0$	$> 0$	$< 0$	$\bar{\beta}$	$\neq 0$	$> 0$	$< 0$
Constant	-1.93	100	0	100	-0.77	77	0	82	0.13	23	23	13
LAUM	-0.10	72	0	77	0.08	46	62	5	0.04	36	46	0
AGE	-0.04	28	0	33	0.07	92	97	0	-0.01	8	3	10
IT	0.00	13	15	0	-0.01	46	0	67	0.00	5	5	3
EH	-0.26	59	5	62	-0.44	85	0	85	0.48	82	85	0
MA	-0.65	54	0	56	-0.53	56	0	67	0.64	72	67	8
RV	0.08	31	26	18	-0.13	23	8	38	0.01	51	31	26
MF	37.81	79	87	0	-31.54	69	0	69	-3.58	13	3	28
PF	1.69	51	46	10	-1.66	38	0	51	0.01	8	5	8
LEV	-0.07	31	10	33	0.00	5	0	10	0.05	23	21	3
HUR	0.10	33	31	10	0.11	3	13	0	-0.11	31	5	31
HWM	0.31	21	28	0	-0.21	26	0	38	0.03	0	3	0
Pseudo $R^2$	96.73				97.17				96.35			

Panel B: Determinants of the overestimation in outperformance and underperformance									
	$\hat{\delta}^+$				$\hat{\delta}^-$				
	$\bar{\beta}$	$\neq 0$	$> 0$	$< 0$	$\bar{\beta}$	$\neq 0$	$> 0$	$< 0$	
Constant	0.36	97	97	0	0.93	100	100	0	
LAUM	0.03	77	85	0	-0.02	82	0	82	
AGE	0.02	56	62	0	-0.05	100	0	100	
IT	0.00	85	0	90	0.00	54	59	0	
EH	0.10	69	69	3	0.10	82	85	3	
MA	0.13	67	54	15	0.14	72	67	10	
RV	-0.02	54	28	33	-0.04	38	5	33	
MF	1.13	5	10	0	1.64	23	28	3	
PF	0.57	64	72	0	0.05	3	8	0	
LEV	0.01	0	5	0	-0.01	0	0	5	
HUR	0.01	0	3	0	0.00	0	0	3	
HWM	-0.01	3	3	3	-0.03	23	0	28	
Pseudo $R^2$	96.56				85.56				

with the equilibrium argument in Berk and Green (2004) that the skilled managers are hired by the funds that charge higher fees.

3.7. Fund characteristics and overestimation in percentile rank-based out- and underperformance

This paper warns against the use of percentile ranks when analyzing outperformance and underperformance. Since they ignore the presence of equal performance, they tend to be inflated. In this section, we refine that result and discuss the fund properties for which we can expect the inflation to be higher. Note first that the overestimation of outperformance in the percentile-rank approach is caused by the large number of instances where the point estimate indicates outperformance while in fact the funds performs equally well or underperforms. It follows that the expected overestimation in outperformance, that is,  $\hat{\delta}_i^+$  in (12), is an increasing function of the value of the equal-performance and underperformance ratio. The higher the latter, the larger is the number of expected false positives. It then follows that  $\hat{\delta}_i^+$  is positively affected by the factors that lead to a higher value of the equal-performance and underperformance ratio. Since a higher value of the equal-performance and underperformance ratio leads to a lower value of the outperformance ratio, it follows also mechanically that variables associated with a higher outperformance ratio tend to lead to a lower value of  $\hat{\delta}_i^+$ . Similarly,  $\hat{\delta}_i^-$  in (13) is positively affected by the factors that lead to a higher value of the equal-performance

and outperformance ratio, and a lower value of the underperformance ratio.

This intuition is confirmed in Panel B of Table 7, where we show the results of the non-linear regression of  $\hat{\delta}_i^+$  and  $\hat{\delta}_i^-$  on the fund characteristics, using the same nonlinear regression model as in (15) except that the link function used for analyzing  $\hat{\delta}_i^+$  and  $\hat{\delta}_i^- \in [-1, 1]$  is  $\tilde{G}(\cdot) \equiv -1 + 2G(\cdot)$  instead of the logistic function  $G(\cdot)$ . Since fund size and fund age lead to a larger underperformance ratio and lower outperformance ratio, and fund size also increases the equal-performance ratio, we find that the percentile rank-based outperformance ratio is inflated for large and old funds. A similar reasoning explains why we find a positive coefficient for most of the Equity Hedge and Macro funds (relatively to Event-Driven funds). For the performance fee, we find that the positive effect on the outperformance ratio dominates and leads to a higher value of  $\hat{\delta}_i^+$ . It thus seems that a higher performance fee is associated with a higher outperformance ratio, but less than the percentile-rank approach would signal.

Similar conclusions hold for the interpretation of the percentile-rank based estimates of the underperformance parameter. As a consequence of the positive effect of fund age and size on the equal-performance ratio and underperformance ratio, and negative effect on the outperformance ratio, we find that for (almost) all samples, the percentile-rank based approach overestimates underperformance for young and small funds. This

corresponds to our previous finding that the underperformance ratio is higher for the old and large funds, implying that the overestimation (i.e., the spread between the rank-based approach and the underperformance ratio) is smaller for these funds. This effect is mitigated in 59% of the samples, where there is a positive interaction effect between fund size and fund age. While these results are indicative of the direction of overestimation in outperformance and underperformance using the percentile rank-based approach, they do not quantify the exact error. Therefore, the bottom line recommendation of the analysis is to account for false discoveries in peer performance and use the proposed triplet of peer performance ratios:  $\hat{\pi}_i^+$ ,  $\hat{\pi}_i^-$ , and  $\hat{\pi}_i^0$ .

**4. Conclusion**

The peer performance of active fund managers is a topic of importance for both academics and practitioners. Inspired by Barras et al. (2010), we argue that the workhorse approach of evaluating peer performance by simply ranking funds on their estimated alpha is prone to suffer from false discoveries: detection of outperformance or underperformance when in reality the funds have the same performance. We introduce a triple-layered peer performance evaluation framework, which, by design, includes the possibility that many fund managers perform equally well as their peers. In our framework, the population of peer funds is segmented into those with truly equal-performance, those that outperform, and those that underperform. We propose a closed-form non-parametric estimator for the corresponding equal-, out-, and underperformance ratios.

We use the proposed peer performance evaluation framework to analyze the peer performance of the Equity Hedge, Event-Driven, Macro, and Relative Value hedge funds in the Hedge Fund Research database over a period ranging from January 2000 to June 2014. The peer group is defined as the set of hedge funds pursuing the same investment style. We show that percentile-rank analyses of out- and underperformance overestimate the outperformance of the funds with a relatively good ranking and overestimate the underperformance of the funds with a worse ranking. An extensive out-of-sample evaluation illustrates the economic value of controlling for false discoveries when using the fund's outperformance ratio rather than the fund's alpha for constructing top quintile investment portfolios. Finally, we validate the proposed methodology also by verifying that the cross-sectional variation in the peer performance ratios is consistent with the life-cycle theory of Berk and Green (2004) and Getmansky (2012).

We thus find the peer performance ratios to be a useful addition to the battery of performance statistics used by investors. By controlling for both relative performance using pairwise tests and adjusting for false discoveries, they provide a unique assessment of the peer performance, which has been shown to be incremental to the existing measures. The proposed peer performance screening plots are useful for both *ex-ante* selection of funds and *ex-post* evaluation of their relative performance. In order to promote its use in practice, we have released the open source statistical package "PeerPerformance" (Ardia and Boudt, 2017) that offers all the functionality documented in this article.

Our research can be extended in various ways. Throughout the paper, we fix the peer group as the funds in the universe with the same investment style. An interesting topic for further research could be to exploit this degree of freedom and determine the definition of the peer group for which the peer performance ratios are most predictive of future fund performance. More generally, the scope of our proposed peer performance framework extends beyond the analysis of hedge fund performance. Additional applications include the selection of trading rules, as in Bajgrowicz and

Scaillet (2012), the analysis of herding among investment professionals, and evaluations of the peer universes used in hedging.

**Acknowledgements**

We are grateful to the Editor, the Associate Editor, two anonymous referees, Marie-Claude Beaulieu, Guido Bolliger, Peter Carl, Philippe Cogneau, Serge Darolles, Geert Dhaene, Ignace De Vos, William Doehler, Michel Dubois, Ivan Guidotti, Lennart Hoogerheide, Simon Keel, Gaëlle Le Fol, Richard Luger, Doug Martin, Philippe Masset, Attilio Meucci, Stefan Nagel, Mikael Petitjean, Gabriel Power, Olivier Scaillet, Enrico Schumann, Jan Tille, Martin Wallmeier, and seminar participants at aeris CAPITAL AG (2011), the R/Finance conference in Chicago (2012), the Computational and Financial Econometrics Conference in London (2013), the 3L conference in Brussels (2013), the Amsterdam-Bonn workshop in Econometrics (2013), the Swiss Economists Abroad conference in St. Gallen (2013), the VU University of Amsterdam (2013), the KU Leuven (2014), the Mathematical Finance Days in Montréal (2014), the Federal Reserve Bank of Saint Louis (2014), Laval University (2015), the 35th International Symposium on Forecasting in Riverside (2015), AFFI (2016), EHL (2016) and Paris Dauphine (2017). Financial support from aeris CAPITAL AG, the Dutch Science Foundation, IFM2 Montréal, and the R/Finance committee (Best Paper Award in 2012) is gratefully acknowledged. The views expressed in this paper are the sole responsibility of the authors. Any remaining errors or shortcomings are those of the authors.

**Appendix A. Bias correction factor**

The bias correction factor is needed because of the truncation, which ensures that  $\hat{\pi}_i^0$  belongs to the feasible space  $[0, 1]$ . Let  $\tilde{\pi}_i^0 \equiv \hat{\pi}_i^0 / c_i^0$  be the estimator without a correction factor:

$$\tilde{\pi}_i^0 \equiv \min \{ \hat{p}_i^0, 1 \},$$

with:

$$\hat{p}_i^0 \equiv \frac{1}{n} \frac{\sum_{j \neq i} I\{\hat{p}_{i-j} \geq \lambda_i\}}{(1 - \lambda_i)}.$$

When there is no estimation error in the *p*-values, i.e.,  $\hat{p}_{i-j} \equiv 1 - F_{i-j}(\hat{\tau}_{i-j})$ , we show that:

$$\mathbb{E}[\tilde{\pi}_i^0] = h(\pi_i^0), \tag{A.1}$$

with:

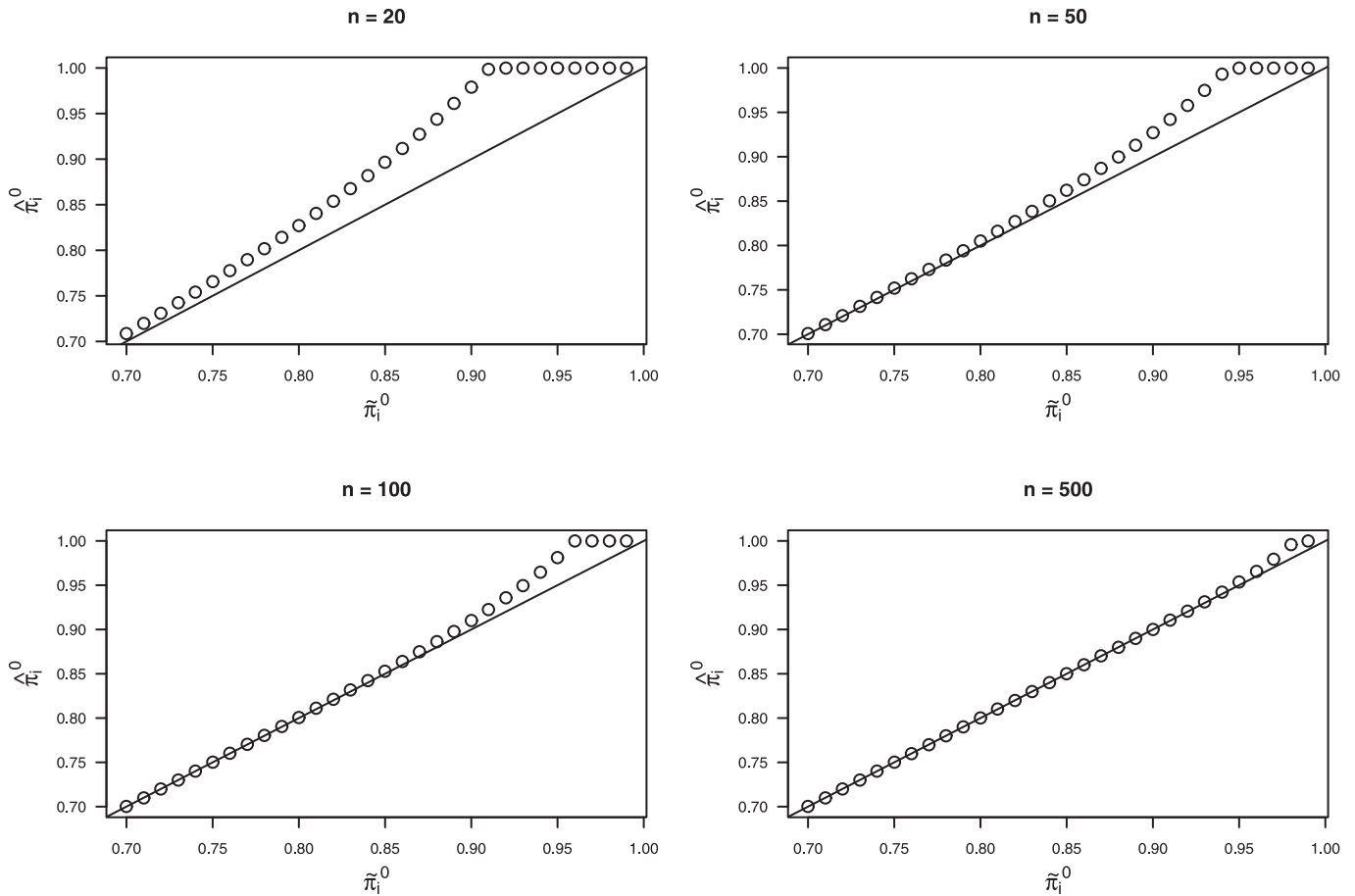
$$h(\pi_i^0) \equiv \pi_i^0 + s_i(-\phi(k) + k(1 - \Phi(k))), \tag{A.2}$$

with  $s_i \equiv \sqrt{\frac{n^{\lambda_i}(n - n^{\lambda_i})}{n^3(1 - \lambda_i)^2}}$ ,  $n_i^{\lambda_i} \equiv (1 - \lambda_i)n_i^0$ ,  $k \equiv \frac{1 - \pi_i^0}{s_i}$ , and where  $\phi$  and  $\Phi$  are the normal density and the cumulative normal distribution, respectively.

Since  $\tilde{\pi}_i^0$  does not estimate  $\pi_i^0$  but  $h(\pi_i^0)$ , we use  $h^{-1}(\tilde{\pi}_i^0)$  to estimate  $\pi_i^0$ . Equivalently, we define the correction factor  $c_i^0$  in (9) as  $c_i^0 \equiv h^{-1}(\tilde{\pi}_i^0) / \tilde{\pi}_i^0$ .

Intuitively, the results in (A.1)–(A.2) are obtained because the randomness in  $\tilde{\pi}_i^0$  comes from drawing the *n* peer funds (with replacement) from a larger population, so we expect  $n\pi_i^0$  equal-performing funds in the sample,  $n\pi_i^+$  underperforming funds, and  $n\pi_i^-$  overperforming funds. Because of the condition (7), all *p*-values exceeding  $\lambda_i$  are equal-performing funds. Let  $n_i^{\lambda_i} \equiv n\pi_i^0(1 - \lambda_i)$  be the expected number of *p*-values above  $\lambda_i$ . Then, the actual number of *p*-values exceeding  $\lambda_i$ , i.e.,  $\sum_{j \neq i} I\{\hat{p}_{i-j} \geq \lambda_i\}$ , follows a binomial distribution with the expected value  $n_i^{\lambda_i}$  and variance  $n_i^{\lambda_i}(n - n_i^{\lambda_i})/n$ , such that:

$$\hat{p}_i^0 \stackrel{d}{\sim} \mathcal{N}\left(\pi_i^0, \frac{n_i^{\lambda_i}(n - n_i^{\lambda_i})}{n^3(1 - \lambda_i)^2}\right).$$



**Fig. A.3.** Impact of the correction factor. This figure presents the plot of  $\tilde{\pi}_i^0$  (horizontal axis; without correction factor) versus  $\hat{\pi}_i^0$  (vertical axis; with correction factor) for various sizes  $n$  of the peer universe.

The expression in (A.2) then follows, since under the location–scale representation of a normal random variable:

$$\mathbb{E}[\tilde{\pi}_i^0] = \pi_i^0 + s_i \mathbb{E}[\min\{Z, k\}],$$

where  $Z$  is a standard normal random variable, and using integration by parts:

$$\begin{aligned} \mathbb{E}[\min\{Z, k\}] &= \int_{-\infty}^k z\phi(z)dz + k(1 - \Phi(k)) \\ &= - \int_{-\infty}^k \phi'(z)dz + k(1 - \Phi(k)) \\ &= -\phi(k) + k(1 - \Phi(k)), \end{aligned}$$

since  $\phi(z)z = -\phi'(z)$ .

Since the non-truncated equal–performance ratio  $\hat{p}_i^0$  is asymptotically normally distributed around the true equal–performance ratio  $\pi_i^0$ , the probability that it exceeds one (and hence the extent of correction needed) increases when  $\pi_i^0$  increases and the variance of the estimate decreases. The latter occurs when the number of peer funds decreases, *ceteris paribus*. We illustrate this by a scatter–plot of the adjusted equal–performance ratio versus the unadjusted one in Fig. A.3 for  $n \in \{20, 50, 100, 500\}$ . Since the truncation leads to an underestimation of the true value, the adjusted equal–performance value is always larger than the unadjusted one and the correction increases when the unadjusted estimate  $\tilde{\pi}_i^0$  is large or  $n$  is small.

### Appendix B. Choice of threshold value $\lambda_i$

Based on Storey (2002), Barras et al. (2010, footnote 10) proposed a bootstrap procedure for determining the value of  $\lambda_i$  in a purely data–driven manner, which minimizes the estimated mean squared error (MSE) of  $\hat{\pi}_i^0$ . Their approach can also be applied to our proposed peer performance analysis. More specifically, we choose  $\lambda_i$  such that an estimate of the MSE of  $\hat{\pi}_i^0(\lambda)$  defined as  $\mathbb{E}[(\hat{\pi}_i^0(\lambda) - \pi_i^0)^2]$  is minimized. First, we compute  $\hat{\pi}_i^0(\lambda)$  using (10) across a range of  $\lambda$  values ( $\lambda \in \{0.3, 0.32, \dots, 0.7\}$ ). Second, for each possible value of  $\lambda$ , we form  $B$  bootstrap replicates of  $\hat{\pi}_i^0(\lambda)$  by drawing with replacement from the  $(n \times 1)$  vector of fund  $p$ -values. These are denoted by  $\hat{\pi}_i^{0*b}(\lambda)$  for  $b = 1, \dots, B$ . Third, we compute the estimated MSE for each possible value of  $\lambda$ :

$$\widehat{\text{MSE}}(\lambda) \equiv \frac{1}{B} \sum_{b=1}^B \left( \hat{\pi}_i^{0*b}(\lambda) - \min_l \hat{\pi}_i^0(l) \right)^2.$$

Finally, we set the optimal  $\lambda_i$  such that  $\lambda_i \equiv \arg \min_{\lambda} \widehat{\text{MSE}}(\lambda)$ . In our empirical application, we use the bootstrap procedure with  $B = 500$  replicates.

### References

Agarwal, V., Naik, N., 2000. Multi-period performance persistence analysis of hedge funds. *J. Financ. Quant. Anal.* 35 (3), 327–342. doi:10.2307/2676207.  
 Aggarwal, R.K., Jorion, P., 2010. The performance of emerging hedge funds and manager. *J. Financ. Econ.* 96 (2), 238–256. doi:10.1016/j.jfineco.2009.12.010.  
 Amihud, Y., Goyenko, R., 2013. Mutual fund’s R2 as predictor of performance. *Rev. Financ. Stud.* 26 (3), 667–694. doi:10.1093/rfs/hhs182.  
 Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59 (3), 817–858. doi:10.2307/2938229.

- Andrews, D., Monahan, J., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 60 (4), 953–966. doi:10.2307/2951574.
- Aragon, G., Liang, B., Park, H., 2013. Onshore and offshore hedge funds: are they twins? *Manage. Sci.* 60 (1), 74–91. doi:10.1287/mnsc.2013.1729.
- Ardia, D., Boudt, K., 2015. Testing equality of modified sharpe ratios. *Finance Res. Lett.* 13, 97–104. doi:10.1016/j.frl.2015.02.008.
- Ardia, D., Boudt, K., 2017. PeerPerformance: luck-corrected peer performance analysis in R. <https://cran.r-project.org/web/packages/PeerPerformance/index.html>.
- Bajgrowicz, P., Scaillet, O., 2012. Technical trading revisited: false discoveries, persistence tests, and transaction costs. *J. Financ. Econ.* 106 (3), 473–491. doi:10.1016/j.jfineco.2012.06.001.
- Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: measuring luck in estimated alphas. *J. Finance* 65 (1), 179–216. doi:10.1111/j.1540-6261.2009.01527.x.
- Berk, J.B., Green, R.C., 2004. Mutual fund flows and performance in rational markets. *J. Polit. Econ.* 112 (6), 1269–1295. doi:10.1086/424739.
- Boyson, N.M., 2010. Implicit incentives and reputational herding by hedge fund managers. *J. Empir. Finance* 17 (3), 283–299. doi:10.1016/j.jempfin.2009.10.005.
- Brown, S.J., Goetzmann, W.N., 1997. Mutual fund styles. *J. Financ. Econ.* 43 (3), 373–399. doi:10.1016/S0304-405X(96)00898-7.
- Carhart, M., 1997. On persistence in mutual fund performance. *J. Finance* 52 (1), 57–82. doi:10.1111/j.1540-6261.1997.tb03808.x.
- Chen, J., Hong, H., Huang, M., Kubik, J.D., 2004. Does fund size erode mutual fund performance? The role of liquidity and organization. *Am. Econ. Rev.* 94 (5), 1276–1302. doi:10.1257/0002828043052277.
- Chen, Y., Cliff, M.T., Zhao, H., 2017. Hedge funds: the good, the bad, and the lucky. *J. Financ. Quant. Anal.* 52 (3), 1081–1109. doi:10.1017/S0022109017000217.
- Cremers, K.M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. *Rev. Financ. Stud.* 22 (9), 3329–3365. doi:10.1093/rfs/hhp057.
- Fama, E., French, K., 2010. Luck versus skill in the cross-section of mutual fund returns. *J. Finance* 65 (5), 1915–1947. doi:10.1111/j.1540-6261.2010.01598.x.
- Ferson, W., Chen, Y., 2015. How many good and bad funds are there, really? Working paper.
- Fung, W., Hsieh, D., 2004. Hedge fund benchmarks: a risk-based approach. *Financ. Anal. J.* 60 (5), 65–80. doi:10.2469/faj.v60.n5.2657.
- Fung, W., Hsieh, D., Naik, N., Ramadorai, R., 2008. Hedge funds: performance, risk, and capital formation. *J. Finance* 63 (4), 1777–1803. doi:10.1111/j.1540-6261.2008.01374.x.
- Getmansky, M., 2012. The life cycle of hedge funds: fund flows, size, competition, and performance. *Q. J. Finance* 2 (1), 1–53. doi:10.1142/S2010139212500036.
- Graham, J., 1999. Herding among investment letters: theory and evidence. *Journal of Finance* 54 (1), 237–268. doi:10.1111/0022-1082.00103.
- Hansen, P.R., 2005. A test for superior predictive ability. *J. Bus. Econ. Stat.* 23 (4), 365–380. doi:10.1198/073500105000000063.
- Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. *Econometrica* 79 (2), 453–497. doi:10.3982/ECTA5771.
- Hunter, D., Kandel, E., Kandel, S., Wermers, R., 2014. Mutual fund performance evaluation with active peer benchmarks. *J. Financ. Econ.* 112 (1), 1–29. doi:10.1016/j.jfineco.2013.12.006.
- Jagannathan, R., Malakhov, A., Novikov, D., 2010. Do hot hands exist among hedge fund managers? An empirical evaluation. *J. Finance* 65 (1), 217–255. doi:10.1111/j.1540-6261.2009.01528.x.
- Joenväärä, J., Kosowski, R., Tolonen, P., 2016. Hedge fund performance: what do we know? Working paper. 10.2139/ssrn.1989410
- Kosowski, R., Naik, N., Teo, M., 2005. Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *J. Financ. Econ.* 84 (1), 229–264. doi:10.1016/j.jfineco.2005.12.009.
- Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the sharpe ratio. *J. Empir. Finance* 15 (5), 850–859. doi:10.1016/j.jempfin.2008.03.002.
- Liang, B., 1999. On the performance of hedge funds. *Financ. Anal. J.* 55 (4), 72–85. doi:10.2469/faj.v55.n4.2287.
- Pástor, L., Stambaugh, R.F., 2002. Mutual fund performance and seemingly unrelated assets. *J. Financ. Econ.* 63 (3), 315–349. doi:10.1016/S0304-405X(02)00064-8.
- R Development Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rosenblatt, M., 1952. Remarks on a multivariate transformation. *Ann. Math. Stat.* 23 (3), 470–472. doi:10.1214/aoms/1177729394.
- Scharfstein, D.S., Stein, J.C., 1990. Herd behavior and investment. *Am. Econ. Rev.* 80 (3), 465–479.
- Storey, J., 2002. A direct approach to false discovery rates. *J. R. Stat. Soc. B* 64 (3), 479–498. doi:10.1111/1467-9868.00346.
- Sun, Z., Wang, A., Zheng, L., 2012. The road less traveled: strategy distinctiveness and hedge fund performance. *Rev. Financ. Stud.* 25 (1), 96–143. doi:10.1093/rfs/hhr092.
- Titman, S., Tiu, C., 2011. Do the best hedge funds hedge? *Rev. Financ. Stud.* 24 (1), 123–168. doi:10.1093/rfs/hhq105.
- White, H., 2000. A reality check for data snooping. *Econometrica* 68 (5), 1097–1126. doi:10.1111/1468-0262.00152.