The peer performance ratios of hedge funds[☆]

David Ardia[a,b,*], Kris Boudt[c,d]

[a] *Institute of Financial Analysis, University of Neuchâtel, Neuchâtel, Switzerland*
[b] *Department of Finance, Insurance and Real Estate, Laval University, Québec City, Canada*
[c] *Solvay Business School, Vrije Universiteit Brussel, Belgium*
[d] *Faculty of Economics and Business, Vrije Universiteit Amsterdam, The Netherlands*

Abstract

We propose to evaluate an investment fund's performance by the percentage of peers the fund outperforms and underperforms, after correction for luck. We call these measures the fund's outperformance ratio and underperformance ratio. When applied to hedge funds, we find that fund size tends to increase the underperformance ratio, but the effect is lower for funds with a longer track record. Our results also indicate that the outperformance ratio is a better predictor of future fund performance than alternative peer performance measures, and that, the best forecast performance is obtained when combining the outperformance ratio with a relative or peer alpha measure.

Keywords: Career hypothesis, decreasing returns to scale, hedge fund, peer performance, performance measurement, relative alpha

[☆] We are grateful to Marie-Claude Beaulieu, Guido Bolliger, Peter Carl, Philippe Cogneau, William Doehler, Geert Dhaene, Ignace De Vos, Michel Dubois, Ivan Guidotti, Lennart Hoogerheide, Simon Keel, Richard Luger, Doug Martin, Attilio Meucci, Stefan Nagel, Mikael Petitjean, Gabriel Power, Olivier Scaillet, Enrico Schumann, 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) and AFFI (2016). 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.

[*] Corresponding author. University of Neuchâtel, Rue A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland. Phone: +41 32 718 1365.

Email addresses: `david.ardia@unine.ch` (David Ardia), `kris.boudt@vub.ac.be` (Kris Boudt)

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, luck must be the driving factor in a frequentist approach.[1] In general, the peer category is defined in a sufficiently broad manner in order to create a level playing field where the peer funds still have ample degrees of freedom to distinguish themselves. Since institutional and retail investors tend to increase their portfolio allocation to outperforming funds (see, *e.g.*, Fung et al., 2008), it is important that the peer performance evaluation criteria employed are reliable signals that the fund managers have a superior ability to yield higher risk-adjusted returns on investment.

The general approach to peer performance evaluation consists of comparing the fund's performance with the performance of its peers. In practice, the most popular approach is to do this in two steps, whereby first the fund's performance is evaluated using standard performance measures, such as the fund's Sharpe ratio. In the second step, the performance of the funds within the peer group is ranked and the percentile ranks are used to classify peer performance as outperformance or underperformance. The main drawback of this estimator is that it ignores the possibility that many funds actually perform equally well. In the extreme case where all funds perform equally well, the percentile-rank rate of outperformance is a random number between zero and one, depending on the luck of the fund. Thus, correcting for luck in peer performance evaluation is of crucial importance.[2]

Recent academic research tends to proceed in one step by directly computing the peer performance measure from the comparison of fund returns with the returns of peer funds. Jagannathan

[1]We refer to Pástor and Stambaugh (2002) for the alternative Bayesian approach where the peer performance of funds corresponds to an analysis of the credible set associated to the posterior distribution of the fund's alpha.

[2]The problem of correcting for luck in peer performance evaluation is different from the correction for luck in the absolute performance evaluation of a fund. Consider for instance the recent paper by Chen et al. (2015) in which the skill of each fund is modelled as a realization from a mixture of J possible normal distributions, and where the fund's skill is estimated as the probability that the fund comes from the highest-skilled group. While this estimate of skill takes the estimation uncertainty into account, the peer analysis requires one to account for the estimation uncertainty in the estimated probabilities of the fund of interest, and the n remaining funds in the peer universe.

et al. (2010) define the fund's *relative alpha* as the intercept obtained in the regression of the hedge fund's returns on benchmark hedge fund investment style returns. Similarly, Hunter et al. (2014) compute the *peer alpha* as the intercept of the Carhart factor model (Carhart, 1997), augmented with a so-called *active peer benchmark* factor. Finally, there are also the *distinctiveness* and *selectivity* peer performance measures proposed by Sun et al. (2012) and Amihud and Goyenko (2013), who define peer performance as the extent of fund performance that cannot be explained by the performance of the peer funds or the factor model, respectively. Like the approach using percentile ranks, all peer performance measures presented above do not correct for the possibility that the significance of the fund's relative performance measure is due to chance.

We solve the issue of adjusting for luck in peer performance analysis by proposing a triple-layered peer performance evaluation framework where 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[3], (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. The corresponding estimators of the peer performance parameters, called *the peer performance ratios*, account for the estimation uncertainty in the underlying risk-adjusted performance measures.[4]

In our application, we focus on the hedge fund industry, where the total assets under management increased to USD 2.90 trillion in the second quarter of 2016 (HFR, 2016). Due to the lack of disclosure requirements, the only public information available to investors is usually the track record of the hedge fund's performance and a general definition of the investment style. Controlling for the investment style is important to distinguish skill from performance due to following

[3]In the sequel, we define the focal fund as the fund of interest, for which the peer performance is measured.

[4]The proposed estimators exploit the properties of the false discovery rate of Storey (2002) and they take an explicit form, as in Barras et al. (2010). An alternative approach for distinguishing true outperformance (respectively underperformance) from luck is to extend the bootstrap methods in Kosowski et al. (2005), Fama and French (2010) and Ferson and Chen (2015), and then to compare the relative performance with "artificially generated" data samples where the variation in fund performance is due entirely to sample variability. However, this simulation framework is computationally more demanding and it does not lead to an explicit form for the peer performance ratios.

similar investment strategies as the peer funds. In fact, Hunter et al. (2014) show that, controlling for funds pursuing similar strategies tends to improve the selection of funds with future outperformance. We therefore define the peer category as the group of hedge funds pursuing the same investment style (*e.g.*, Macro funds). A hedge fund is considered to be talented if it outperforms most of its peers. We demonstrate how this analysis can be conducted using the proposed peer performance ratios.

The investment universe considered in this study is comprised of 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. It follows that the resulting peer performance ratios should not be interpreted as a measure of skill, but instead as a reliable signal related to the capacity of the fund to generate relatively higher risk-adjusted returns (in excess of expenses) compared with the other funds in the peer category (see Berk and Van Binsbergen, 2015). 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.

As expected, we show that for the majority of hedge funds, the return history indicates that we cannot exclude that they perform equally well as most of their peers. Thus, on average, the standard approach for estimating peer performance using percentile ranks is too optimistic about the outperformance of the funds with a relatively good ranking and too pessimistic about the underperformance of the funds with a worse ranking. The correction for luck is economically relevant since the average downward correction of outperformance and underperformance is around 34%. A first practical implication of finding that the track record of past net returns is less informative than expected is the high importance of seeking for diversification when investing in hedge funds. Secondly, while most of the funds perform equally well as the majority of their peers, there is still a substantial heterogeneity in the out- and underperformance ratios, implying their potential usefulness for peer fund selection. We analyze the determinants of peer performance and the information value for predicting fund returns.

A cross-sectional analysis of our sample of peer performance measures confirms theoretical predictions regarding the impact of fund size on the peer performance of hedge funds. In particular, we find that when a fund has more assets under management, the outperformance ratio tends to be lower and the underperformance ratio is higher. This result is intuitive, since, as predicted by the liquidity hypothesis of Chen et al. (2004), large funds have fewer opportunities to deploy their talent, especially when investing in less liquid assets. We further investigate the interaction between the effects of fund size and fund age on peer performance and show that the positive impact of fund size on the underperformance ratio shrinks with fund age. Our empirical evidence that the old and large funds tend to underperform less than the young and large funds is consistent with the career and reputation hypothesis, which argues that mature funds tend to herd more in order to avoid a loss in reputation (see, *e.g.*, Scharfstein and Stein, 1990; Brown et al., 2001).

Finally, we address the important question of the informativeness of peer performance when predicting future performance. Using portfolio sorts, we first show that a quarterly rebalanced portfolio investing in the top quintile hedge funds in terms of the outperformance ratio significantly outperforms the same investment based on alternative (peer) performance measures such as the relative alpha (Jagannathan et al., 2010), the peer alpha (Hunter et al., 2014), the distinctiveness (Sun et al., 2012) and the selectivity (Amihud and Goyenko, 2013). Importantly, we find that the outperformance ratio and relative (or peer) alpha are complementary. In fact, the highest risk-adjusted performance is obtained when a sequential approach is used in which first the top forty percent hedge funds is selected based on the magnitude of relative performance, as measured by the fund's relative alpha or peer alpha, and then the top half is selected in terms of the highest outperformance ratio. The same result is obtained when using value-weighted portfolios, indicating that the result is not driven by small funds which may not be investable. We also use multivariate regressions to analyze the predictive power of the outperformance ratio for the next quarter's fund return. Controlling for a host of other influences such as the first order autocorrelation in fund returns, the return volatility, the fund style, age, size, fee structure or capital inflow, we confirm the two main findings of the portfolio sorting analysis: The outperformance ratio is a good predictor of the future hedge fund return and the best forecast performance is obtained by exploiting the complementarity in the signal captured by the values of the outperformance ratio and the relative or peer alpha.

Altogether, the empirical results strongly support our view that the proposed peer performance ratios are useful for *ex-post* performance evaluation of hedge funds and *ex-ante* screening of the universe in order to select potential investments for a deeper, more fundamental, analysis of the fund's investment style, portfolio holdings, expenses and organization.

The remainder of this article is organized as follows. In Section 2, we summarize previous studies of peer performance analysis and we define our research hypotheses. Section 3 introduces the methodology employed to obtain the peer performance ratio. Section 4 presents our main results regarding the importance of luck-adjustment for peer performance analysis, as well as explaining the effects and determinants of peer performance for hedge funds.[5] In Section 5, we give our conclusions.

2. Literature review and research hypotheses

In this literature review, we first discuss the state of the art in terms of estimating peer performance and how the proposed peer performance methodology fills a research gap. We then formulate the hypotheses about the peer performance of hedge funds that we test in Section 4 using the net returns of hedge funds in the Hedge Fund Research database over the period 2000–2014.

[5] All of the computations performed in this study employed the R statistical computing language (R Development Core Team, 2013) with the package **PeerPerformance** (Ardia and Boudt, 2016), which is freely available at `https://github.com/ArdiaD/PeerPerformance`.

2.1. Measures of peer performance

In academic research it is common to measure the talent of a (hedge) fund manager by regressing the fund returns (in excess of the risk-free rate) on the four Carhart factors, or the seven Fung and Hsieh hedge fund risk factors (Carhart, 1997; Fung and Hsieh, 2004).[6] The intercept (usually called the *alpha*) of that regression is interpreted as a measure of talent. When the estimated alpha is significantly different from zero, the fund is classified as talented. While this binary outcome is a relevant solution to a single fund evaluation problem, it does not solve the question of peer performance faced commonly by practitioners. Their investment problem is not to invest (or not) in a single fund or risk factor, but instead they must select funds from a universe of peer funds or evaluate a fund with respect to peer funds. Moreover, the benchmark in the regression is often not investable, whereas the peer universe of funds is constructed so that this is the case.

The general solution to this problem is to explicitly incorporate the performance of the peer funds in the evaluation. Jagannathan et al. (2010) define the fund's *relative alpha* as the intercept obtained in the regression of the hedge fund's returns on benchmark hedge fund investment style returns. Their empirical evidence supports the claim that this relative alpha measure allows one to identify the managers who are likely to have superior skills relative to their peers. Hunter et al. (2014) combine the classical factor model approach with the relative alpha idea in Jagannathan et al. (2010) by proposing to include a so-called *active peer benchmark* factor in the risk factor model of Carhart (1997) and Fung and Hsieh (2004). In the remainder of the paper, we call this the *peer alpha*. This expanded factor model is a special case of Pástor and Stambaugh (2002) who show that including additional factors, even when they are not priced, can increase the efficiency

[6] For practitioners, the best-known definition of peer performance is probably the Morningstar rating, which assigns five stars to the 10% best performing funds, one star to the 10% worst performing funds, and three stars to the middle 35% performing funds. The other segments (*i.e.*, 22.5% each) receive four and two stars. We argue that this "percentile-rank based" approach to peer fund analysis can be considered *naive* because it completely ignores the possibility that all funds perform equally well, in which case they should all receive the same number of stars. The practical implementation of the Morningstar rating methodology implies that the funds with the most luck in the sample of equal-performing funds will receive the highest rating, and thus the noise during estimation drives the performance analysis.

of the estimate of a fund's talent.

Another solution consists of associating outperformance to funds for which the track record or returns cannot be explained by their peer returns or the factor model. Sun et al. (2012) recommend to do fund selection based on the fund's *distinctiveness*, defined as one minus the correlation between the fund's return and the average return of the funds in the peer universe. The fund's distinctiveness is thus a peer performance measure, which does not explicitly take the sign of the peer performance into account. Funds with unexpected (positive or negative) returns with respect to the returns of the peer group or a factor model are predicted to be skilled and therefore outperforming funds. This approach is related to the model of Titman and Tiu (2011) who show that, for a Sharpe ratio maximizing fund invested in a risk-free asset, a publicly available index, and a proprietary strategy, the R-squared of the regression of the investor's excess returns on systematic factors is inversely related to the fund's Sharpe ratio and to its information ratio. Amihud and Goyenko (2013) propose to exploit these results and recommend picking funds based on their *selectivity*, defined as one minus the R-squared of the regression estimating the fund's alpha.

In this study, we introduce a new peer performance framework, which measures the peer performance of a focal fund i belonging to a peer universe of $n+1$ funds using three peer performance parameters: (i) π_i^0: the proportion of funds in the peer group that perform equally well as fund i, (ii) π_i^+: the proportion of funds in the peer group that are outperformed by fund i, and (iii) π_i^- the proportion of funds in the peer group that outperform fund i. The major strengths of the proposed peer performance ratios are both that relative performance between two funds must cross a threshold of statistical significance in order to be counted as evidence of difference in performance, and that we use the false discovery rate methodology to obtain peer performance estimates that are robust to false positives.

2.2. *Hypotheses on the peer performance of hedge funds*

The proposed framework of assessing the peer performance based on under-, equal-, and outperformance ratios allows us to shed new light on the determinants and consequences of peer performance by hedge funds. Such an analysis is of the utmost importance for the functioning of financial markets, since defendants of market efficiency tend to praise competition between hedge funds for detecting arbitrage opportunities and exploiting them almost instantaneously. The virtuous effects of hedge fund competition on stock market efficiency would disappear if hedge fund managers ignore their skills and copy each other's behavior. This tendency to herd would be apparent in our analysis in the form of high values of the equal-performance ratio.

Throughout the empirical analysis, four working hypotheses will be tested using a combination of nonlinear multivariate regressions and portfolio sorts. The hypotheses focus on the effect of fund size and fund age on peer performance, on the one hand, and the predictive power of the outperformance ratio to forecast future hedge fund returns, on the other hand. The sample consists of the net returns of hedge funds in the Hedge Fund Research database over the period 2000–2014. To account for the time-variation in the (peer) performance of hedge funds, the empirical tests of the hypotheses use rolling peer performance evaluation windows.

Hypothesis I: Ceteris paribus, fund size deteriorates the peer performance of hedge funds, i.e.*, it decreases the outperformance ratio and increases the underperformance ratio.*

Hypothesis I is expected in the presence of decreasing returns to scale in active management. Chen et al. (2004) attribute the diseconomy of scale to the fact that small hedge funds have more opportunities to deploy their talent, especially when investing in less liquid assets. In the model of Berk and Green (2004), the diseconomy of scale is compensated by the fact that bigger funds attract higher-skilled managers. In independent and concurrent work, Pástor et al. (2015) provide empirical evidence against this equilibrium outcome and, in line with our hypothesis, they find that the younger and smaller funds tend to outperform the older and larger funds, because the new entrants tend to be more skilled. This is also consistent with Fung et al. (2008), who show that during the 1995–2004 period, alpha-producing hedge funds experienced far greater and steadier capital inflows than their less fortunate counterparts. Fung et al. (2008) conclude that these capital inflows attenuated their ability to continue to deliver alpha in the future.

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 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.

Hypothesis II: Ceteris paribus, large hedge funds with a short track record tend to underperform more than their peers with equal fund sizes but a longer track record.

Our analysis of the effects of peer performance on fund performance focuses on the use of the outperformance ratio as a predictor for future fund performance.[7] We are not interested in the behavioral interpretation of the relationship, but in the existence of forecasting power. We expect the outperformance ratio of a hedge fund to be useful for hedge fund selection and for predicting future fund performance. Compared with the distinctiveness and selectivity measures proposed by Sun et al. (2012) and Amihud and Goyenko (2013), the proposed outperformance ratio has the advantage of measuring simultaneously the fund's specialness with respect to its peers and the success of the fund in translating this specialness into alpha. Compared with the relative alpha and peer alpha measure of Jagannathan et al. (2010) and Hunter et al. (2014), the

[7]An interesting direction for further research is to investigate the effects of the underperformance ratio on the survival of the fund and on potential changes in the fund investment style.

proposed outperformance ratio has the advantage of controlling for luck in the peer performance evaluation. In that sense, the peer performance ratios are complementary to the existing methods and we expect the highest gains in performance when combining the two approaches. We study this hypothesis for a quarterly updated portfolio and thus focus on the short-run (*i.e.*, one quarter ahead) information signal of peer performance for the future performance of hedge fund managers.

Hypothesis IIIa: A quarterly updated fund of hedge funds, equally invested in the top quintile funds in terms of alpha, relative alpha, peer alpha, distinctiveness or selectivity, underperforms in terms of risk-adjusted performance, the quarterly updated fund of funds investing in the top quintile funds based on the outperformance ratio.

Hypothesis IIIb: A sequential approach of first selecting the top 40% funds based on alpha, relative alpha, peer alpha, distinctiveness or selectivity and then selecting the top half in terms of outperformance ratio yields a higher risk-adjusted performance than the single-indicator approach of investing in the top quintile funds in terms of alpha, relative alpha, peer alpha, distinctiveness, selectivity or the outperformance ratio.

Hypotheses IIIa and *IIIb* focus on the economic value of the outperformance ratio for fund selection. They are evaluated by means of an investment analysis. This approach does not control for the other possible influences on the quarterly fund performance. Our final hypothesis is that the outperformance ratio remains a significant predictor of the quarterly return, after controlling for other influences of hedge fund returns, such as the fund's lagged return performance, the fund's size, fee structure and age, the period fixed effects and the fund return volatility.

Hypothesis IV: Ceteris paribus, the higher the outperformance ratio of a hedge fund, the higher is, on average, the subsequent short-term return.

3. Estimation of peer performance ratios

3.1. Definitions and estimation strategy

We consider a universe with a total of $n+1$ funds. We denote Δ_{i-j} as the (true) difference in performance between fund i and j $(i \neq j)$, and $\widehat{\Delta}_{i-j}$ as the corresponding estimate. Throughout this study, we use the *hat* superscript symbol to denote sample-based estimates.

We want to estimate the percentage of funds that have equal, less, or greater risk-adjusted performance than fund i. Denote π_i^0 (n_i^0), π_i^+ (n_i^+), and π_i^- (n_i^-) as the proportion (number) of funds where $\Delta_{i-j} = 0$, $\Delta_{i-j} > 0$, and $\Delta_{i-j} < 0$, respectively.

A simple but biased estimator of the outperformance ratio π_i^+ is the percentage of funds for which the performance of the focal fund is *significantly* higher, such as at the 10% level.[8] As a solution, it might be tempting to use a sequence of multiple hypothesis tests, such as *Hotelling's T2* test of equality of Sharpe ratios in Pav (2014), with family-wise error rate control of Type I errors over the sequence of tests.[9] 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.

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 sam-

[8]This measure would not be acceptable because it can be biased for two reasons. First, because of multiple testing based on a large number of peer funds, under the null of no significant difference in talent, we expect to find that the average estimated percentage of outperformance is 10%, whereas it is actually is 0%. Second, if the focal fund is truly outperforming its peers, it might be due to bad luck that the test statistic lies in the region of non-rejection.

[9]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).

ple of p-values $\widehat{p}_{i-j}$ associated with a two-sided test of the null hypothesis $H_0 : \Delta_{i-j} = 0$, for $j = 1, \ldots, n$, $j \neq i$, 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.[10] 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 to 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, so 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.

3.2. Test for equal performance of two funds

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 seven Fung and Hsieh hedge fund risk factors (Carhart, 1997; Fung and Hsieh, 2004).[11] Let $\mathbf{f}_t$

[10]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.

[11]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).

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 regression of $(r_{i,t} - r_{j,t})$ on $\mathbf{f}_t$:

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

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

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

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

We denote $\widehat{\tau}_{i-j} \equiv \widehat{\Delta}_{i-j}/\widehat{\mathrm{se}}_{i-j}$ 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.[12]

[12]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.

3.3. The equal-performance ratio

A crucial feature of the proposed estimators is the difference in the distribution of the p-values $\widehat{p}_{i-j}$ when $\Delta_{i-j} = 0$ versus $\Delta_{i-j} \neq 0$.

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

$$(A1): \quad \mathbb{P}[\widehat{p}_{i-j} < \lambda_i \,|\, \Delta_{i-j} \neq 0] = 1\,. \tag{2}$$

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

In the case of equal performance, and provided that the estimated $\widehat{F}_{i-j}$ 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 $\widehat{p}_{i-j}$ exceeding λ_i when $\Delta_{i-j} = 0$ is $1 - \lambda_i$:

$$(A2): \quad \mathbb{P}[\widehat{p}_{i-j} \geq \lambda_i \,|\, \Delta_{i-j} = 0] = 1 - \lambda_i\,. \tag{3}$$

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 λ_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:

$$\mathbb{E}\left[\sum_{j\neq i} I\{\widehat{p}_{i-j} \geq \lambda_i\}\right] = \underbrace{\sum_{\substack{j\neq i \\ \Delta_{i-j}=0}} \mathbb{E}\left[I\{\widehat{p}_{i-j} \geq \lambda_i\}\right]}_{=(1-\lambda_i)\, n_i^0} + \underbrace{\sum_{\substack{j\neq i \\ \Delta_{i-j}\neq 0}} \mathbb{E}\left[I\{\widehat{p}_{i-j} \geq \lambda_i\}\right]}_{=0}$$
$$= (1-\lambda_i)\, n_i^0\,,$$

where $I\{\cdot\}$ denotes the indicator function, which equals one if the condition holds and zero otherwise. Hence, a natural estimator for n_i^0 is the number of estimated p-values exceeding λ_i divided by $1-\lambda_i$:

$$\widehat{n}_i^0 \equiv c_i^0 \min\left\{\frac{\sum_{j\neq i} I\{\widehat{p}_{i-j} \geq \lambda_i\}}{1-\lambda_i}, n\right\}\,, \tag{4}$$

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:

$$\widehat{\pi}_i^0 \equiv \frac{\widehat{n}_i^0}{n}\,. \tag{5}$$

Since the estimator $\widehat{\pi}_i^0$ has the form of a sample average, we expect that in most relevant cases, the estimator $\widehat{\pi}_i^0$ is not only unbiased but also consistent for π_i^0. The proof of this requires a suitable law of large numbers that allows for the dependence in the p-values.[13]

The practical computation of the equal-performance ratio thus requires us to choose the threshold value λ_i. The larger the value of λ_i, the more likely it is that assumptions (A1) and (A2) are

[13] 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 tests $\Delta_{i-j} = 0$, 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 lucky fund and a talented fund. See Bajgrowicz and Scaillet (2012, Appendix F) and the references therein.

satisfied, but the fewer observations enter the summation used to estimate the equal-performance ratio in (4)–(5). We address this bias-variance trade-off in the estimation of π_i^0 by optimizing the choice of λ_i based on a data-driven approach to determine the value of $\lambda_i \in \{0.3, 0.32, \ldots, 0.7\}$ which minimizes the estimated mean squared estimation error of $\widehat{\pi}_i^0$. See Appendix B for more details. In our application to hedge funds, λ_i varies between 0.3 and 0.7 with a median (resp. mean) value of 0.66 (0.58). In most cases, the concern of avoiding bias thus dominates the objective of precision in estimating the equal-performance ratios.

3.4. The out- and underperformance ratios

Given the estimate of the number of peer funds with equal performance $\widehat{n}_i^0$, and the observed performance differences $\widehat{\Delta}_{i-j}$, 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 $\widehat{\tau}_{i-j} \equiv \widehat{\Delta}_{i-j}/\widehat{\mathbf{se}}_{i-j}$) 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 $\sum_{j \neq i} I\{\widehat{\Delta}_{i-j} \geq 0\} \geq n/2$), we first estimate n_i^+, and then attribute $n - \widehat{n}_i^0 - \widehat{n}_i^+$ to $\widehat{n}_i^-$. The estimate $\widehat{n}_i^+$ 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 β^+, *i.e.*, the number of funds for which the estimated test statistic $\widehat{\tau}_{i-j}$ exceeds the (estimated) β^+-quantile of the distribution of $\widehat{\tau}_{i-j}$ under the null hypothesis, which is denoted by $\widehat{q}_{i-j}^{\beta^+}$)[14], adjusted for the false positives (*i.e.*, cases where $\widehat{\tau}_{i-j} > \widehat{q}_{i-j}^{\beta^+}$ when $\Delta_{i-j} \leq 0$) and false negatives (*i.e.*, cases where

[14]In our application we use the asymptotic normal distribution of $\widehat{\tau}_{i-j}$ and thus we set $\widehat{q}_{i-j}^{\beta^+}$ as the β^+-quantile of the standard normal distribution.

$\widehat{\tau}_{i-j} < \widehat{q}_{i-j}^{\beta^+}$ when $\Delta_{i-j} > 0$):

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

The approximation error in the last step of (6) is small for well-chosen values of β^+. First, given $\widehat{n}_i^0$, we can infer that the number of false positives is $\widehat{n}_i^0(1-\beta^+)$ if $\Delta_{i-j} = 0$. $\widehat{n}_i^0$ is an unbiased estimate of n_i^0, so $\widehat{n}_i^0(1-\beta^+)$ is an unbiased estimate for the false discoveries when $\Delta_{i-j} = 0$. Since there are relatively more cases where $\widehat{\Delta}_{i-j}$ 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 β^+ as sufficiently low.[15]

In the opposite case where there are more point estimates of underperformance by fund i (*i.e.*, when $\sum_{j\neq i} I\{\widehat{\Delta}_{i-j} \geq 0\} < n/2$), almost all of the peer funds either perform equally well or outperform fund j, such that n_i^- is accurately estimated by $\sum_{j\neq i} I\{\widehat{\tau}_{i-j} \leq \widehat{q}_{i-j}^{\beta^-}\}$ where we need

[15] For the estimation of the number of good and bad managers, Barras et al. (2010) follow the same reasoning, which was later criticized by Ferson and Chen (2015). We agree with Ferson and Chen (2015) that, when for many peer funds j, the difference in performance Δ_{i-j} is not zero but small, there is a tendency to overestimate the equal-performance ratio. Combined with the overestimation of π_i^0 when Δ_{i-j} is statistically small, our implementation sets β^+ to a high value (*i.e.*, 40%), and therefore the instances of confusion between a case of outperformance with underperformance based on the test statistic $\widehat{\tau}_{i-j}$ are expected to be small. Importantly, we do not worry much about the overestimation of π_i^0 when $\Delta_{i-j} = 0$, because, economically speaking, the overestimation of π_i^0 and the likely underestimation of both π_i^+ and π_i^- is acceptable, as it reflects the simple fact that the data is not sufficiently informative to distinguish equal-performance from out- or 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.

to set β^- as sufficiently high so that the number of false negatives becomes negligible. In the application, we set $\beta^+ = 0.4$ to avoid the false negatives and $\beta^- = 0.6$ 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:

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

and:

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

4. Hedge fund peer performance

We now illustrate the practical relevance of the proposed peer performance measures to analyze the performance in hedge funds. The analysis is centered around three main questions. How much of the traditional percentile rank-based outperformance is due to luck? Is there empirical support in the hedge fund return data for the liquidity hypothesis that larger funds tend to outperform less? Are improvements in peer performance a signal of better performance by a hedge fund? These are important questions that the proposed peer performance framework allows us to answer. We address these questions based on the universe of active and dead U.S. hedge funds included in the Hedge Fund Research (HFR) database as of July 2014.[16]

[16]Overall, our database contains 15,370 funds. For each of the (rolling) quarterly database update, 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 styles. We further delete incoherent and duplicate entries following the approach described in Joenväärä et al. (2016). On average, there are 1,391 funds per (rolling) quarterly samples, with a minimum of 921 funds and a maximum of 1,634 funds.

For the calculation of the peer performance measure, we need to decide on the peer group to use and the measure of performance used to compare the performance of the hedge fund with its peers. In the main analysis, we define the peer group as the set of hedge funds following the same investment style (Equity Hedge, Event-Driven, Macro, and Relative Value). In the robustness analysis, we compare the results obtained when computing the peer performance measures with respect to all peer funds. We find that, while the main results are qualitatively similar using all funds as the peer groups, the significance of the results and the predictive power of the peer performance measures in forecasting fund performance is higher when restricting the peer group to the funds pursuing the same investment style. This is also consistent with Brown and Goetzmann (2003), who show that fund style is a major determinant of the cross-sectional variability in fund performance. For this reason, we recommend to use the funds pursuing the same investment style as a peer group and present this approach as our main analysis.[17]

Another decision to make when computing the peer performance ratios is the choice of performance measure. Throughout the paper, 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 seven factors in Fung and Hsieh (2004) factor model in (1). The p-values are computed using heteroscedasticity and autocorrelation robust standard error estimators (Andrews, 1991; Andrews and Monahan, 1992).

We include the seven Fung and Hsieh (2004) factors because they are generally accepted as the main risk factors to describe the returns of a balanced hedge fund. The sample of hedge funds that we consider also includes a large proportion of hedge funds that are invested only in equities.

[17] An alternative of defining peers based on the declared investment style would be to 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 obtains the peer funds through cluster analysis of the returns 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. The robustness of our results to the choice of peer group is examined throughout this empirical analysis.

As such, we choose to combine the seven Fung and Hsieh (2004) factors (which do not include the equities' high minus low value factor and the equities' momentum factor) with the equity risk factors proposed by Carhart (1997). This leads us to estimate the hedge funds' alpha based on nine risk factors, namely: the US aggregate market factor, the small minus big factor, the high minus low factor and the momentum factor from the website of Kenneth French as equity risk factors, the monthly change in the ten-year treasury constant maturity yield is used as an aggregate bond-oriented risk factor, and the monthly change in the difference between Moody's Baa yield and the ten-year treasury constant maturity yield is used as a credit risk factor in bond markets. Finally, to capture the risk related to trend-following investments, we include the bond, currency, and commodity trend-following risk factors.[18]

In this section, we first present the data. We then investigate the average cross-sectional distribution of peer performance and we quantify how significant the adjustment for luck is in peer performance ratios compared with the traditional percentile-rank approach. Our main empirical results concern the determinants and effects of peer performance, where we use regression analysis to study the fund characteristics that determine the peer performance and to analyze the effects of peer performance on future fund performance.

4.1. Data description

To account for the time-variation in the hedge funds' alpha, the peer performance ratios are computed on 39 (rolling) samples of five years (*i.e.*, 60 monthly observations) of net returns over the period of 2000–2014.[19] Each sample starts at the beginning of the quarter and is indexed by $q = 1, \ldots, 39$ (*i.e.*, January 2000–December 2004, April 2000–March 2005,..., July 2009–June

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

[19]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.

2014). In order to avoid survivorship bias, the fund composition of each sample changes, and tracks the funds available in the HFR active and dead databases for the corresponding period. Because the universe changes, the longitudinal time series of peer performance ratio estimates ($\widehat{\pi}^{+}_{i,q}$, $\widehat{\pi}^{-}_{i,q}$ and $\widehat{\pi}^{0}_{i,q}$) is thus unbalanced.

The input data for our analysis comprise the fund net returns and the fund characteristics, which are available in the HFR database. In Table 1 we present the corresponding summary statistics 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. The remaining funds belong to the categories Macro (21%), Relative Value (16%) and Event-Driven (9.8%). 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. The average annual return is around 8.5% per year, with 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 is asymmetric. On average, we find a positive skewness with an average minimum performance of –35.25% and an average top performance of 76.82%. 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 9.65% and average range of 104.58%) than for the Event-Driven hedge funds (average standard deviation of 7.85% and average range of 65.73%).

[Insert Table 1 about here]

The distribution of net returns needs to be adjusted for the risk factor exposure prior to testing for equal performance. Panel B reports the distribution of the average (annualized) monthly alpha of the hedge funds obtained by ordinary least squares estimation of the linear model regressing the net returns in excess of the risk-free rate against the nine Fung and Hsieh (2004)-Carhart (1997) 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 alpha (5.86%), followed by the Relative Value funds (5.03%), and Event-Driven funds (3.94%). The worst performance in terms of average alpha is by the Equity Hedge investment funds (3.08%). 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, age, 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.

For all hedge funds, we analyze not only the fund's alpha and the proposed peer performance ratios, but benchmark the outperformance ratio against alternative approaches to evaluate the fund's peer performance, namely the fund's relative alpha (Jagannathan et al., 2010), peer alpha (Hunter et al., 2014), distinctiveness (Sun et al., 2012) and selectivity (Amihud and Goyenko, 2013) measures. Like the peer performance measures, we compute these relative performance measures on rolling samples of five years of monthly net returns. The alpha measure (*i.e.*, $\widehat{\alpha}$) is the intercept of the linear regression estimated using the nine risk factors constituting the union of the four factors in Carhart (1997) and the seven factors in Fung and Hsieh (2004). For the estimation of the relative alpha, denoted by $\widehat{\alpha}^{\text{rel}}$, we follow the methodology of Jagannathan et al. (2010). We estimate the intercept of a linear regression model with 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.[20] For the peer

[20]Overall, we use the 37 HFRI style factors corresponding to the HFR styles described in `https:`

alpha measure, denoted by $\widehat{\alpha}^{\text{peer}}$, we follow Hunter et al. (2014), and estimate the peer performance measure as the intercept of a regression on the four factors of Carhart (1997) and the self-reported style factor.[21] The distinctiveness, denoted by $\widehat{D}$, is measured as the correlation between the fund's returns and the self-reported style factor, while the selectivity, denoted by $\widehat{S}$, is defined as one minus the R-squared of the regression used to compute $\widehat{\alpha}$. For each performance measure, we have 46,378 estimates. Since we have 39 estimation dates in our database, this corresponds to, on average, almost 1,200 funds per estimation window.

Table 2 reports the Pearson and Spearman correlations between the various (peer) performance measures. We see that there is a high correlation between the fund's alpha and the outperformance ratio (around 76%). The correlation between the fund's relative and peer alphas and the outperformance ratio is also high. The correlation is lower for the fund's distinctiveness and selectivity (a Pearson correlation of around 12% and 25%, respectively). This indicates that, even though the different measures claim a commonality in what they proxy (*i.e.*, the fund's talent), they are complementary in terms of what exactly is captured.

[Insert Table 2 about here]

4.2. 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 Figure 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 in 50 equal-sized buckets, where "Bucket 1" corresponds to the best performing funds in terms of alpha, while in the right part, a barplot displays the average estimated

//www.hedgefundresearch.com/hfr-hedge-fund-strategy-classification-system. The self-reported style factor is computed as an equally-weighted average of returns for all hedge funds with the same strategy.

[21] We found qualitatively similar results when using AIC instead of BIC criterion or when using value weighted averages instead of equal weighted averages when computing the self-reported style factor.

out-, equal-, and underperformance 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 see for instance that a decrease of the fund's alpha by 10 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.

[Insert Figure 1 about here]

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 underperformance ratios for the fund mentioned in the row compared with other funds with the given investment style indicated in the column header.

We can see that there is substantial scope for out- or underperformance within an investment style. For instance, the 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. Thus, during the analysis of the determinants of hedge fund performance, it is important to control for the hedge fund style to explain the individual drivers of peer performance.

[Insert Table 3 about here]

4.3. Correction for luck

In the introduction, we position the proposed peer performance ratio framework against the standard percentile-rank approach for estimating out- and underperformance, where the proposed methods have the advantage of considering luck. A natural question that we now investigate is the order of magnitude of the correction for luck.

Under the percentile-rank approach, a fund with rank k outperforms $n - k$ out of n peers. Due to the high frequency of equal-performance in hedge funds, on average, this ranking is too optimistic and it needs to be corrected for luck. Thus, we measure the correction for luck for hedge fund i based on the difference between the (luck-adjusted) outperformance ratio and the percentile-rank estimate:

$$\widehat{\delta}_i^+ \equiv \underbrace{\text{luck-adjusted outperformance}}_{\equiv \widehat{\pi}_i^+} - \underbrace{\text{outperformance based on ranks}}_{\equiv \left(\frac{n-k+1}{n}\right)} .$$

Similarly, the correction for (bad) luck regarding the underperformance ratio is:

$$\widehat{\delta}_i^- \equiv \underbrace{\text{luck-adjusted underperformance}}_{\equiv \widehat{\pi}_i^-} - \underbrace{\text{underperformance based on ranks}}_{\equiv \left(\frac{k-1}{n}\right)} .$$

Both correction terms can take values in the range of –100% to 100%. For both out- and underperformance, we expect mostly negative corrections when considering the fact that many of the so-called outperforming (respectively underperforming) funds based on the percentile-rank approach are *lucky* (respectively *unlucky*) and are actually performing equally as well as many of their peers.

The left plot in Figure 2 presents the (average) correction for luck in outperformance for the 50 buckets of hedge funds in our universe, which are ranked by decreasing alpha. As expected, the correction for luck is always negative: The percentile-rank approach overestimates outperformance by an average of 34.1%. Similarly, in the right plot, we can see that the percentile-rank approach significantly overestimates underperformance by an average of 34.8%.

[Insert Figure 2 about here]

4.4. Testing Hypotheses I and II: Effect of fund size and fund age on peer performance

Based on the presence of decreasing returns to scale in active management, we hypothesize in Section 2 that the relatively larger funds in the peer group tend to underperform with respect to their relatively smaller peers. This positive effect of fund size on the underperformance ratio is expected to be smaller for funds with a longer track record, which tend to have a higher reputation risk and are thus expected to herd more.

We test these hypotheses in our longitudinal time series of quarterly updated peer performance ratios $\widehat{\pi}^{+}_{i,q}$, $\widehat{\pi}^{-}_{i,q}$ and $\widehat{\pi}^{0}_{i,q}$, with $q = 1, \ldots, 39$. The determinants of the peer performance ratios are analyzed using a nonlinear regression framework where 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:

$$\widehat{\pi}^{\bullet}_{i,q} = G\big(\beta_{0,q} + \beta_{1,q}LAUM_{i,q-1} + \beta_{2,q}AGE_{i,q-1} + \beta_{3,q}IT_{i,q-1} + \boldsymbol{\gamma}'_q \boldsymbol{CTRL}_i\big) + \varepsilon_{i,q}\,, \qquad (7)$$

where $G(\cdot)$ is the logistic function, $\boldsymbol{\gamma}_q$ is a (8×1) vector of parameters for the control variables and $\varepsilon_{i,q}$ is the error term. The control variables are:

$$\boldsymbol{CTRL}_i \equiv \big(EH_i, MA_i, RV_i, MF_i, PF_i, LEV_i, HUR_i, HWM_i\big)'\,,$$

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).[22] MF_i and PF_i are the fund's man-

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

agement 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). Summary statistics are reported in Panel C of Table 1.

The dependent variable $\widehat{\pi}^{\bullet}_{i,q}$ is either the outperformance ratio $\widehat{\pi}^{+}_{i,q}$, the underperformance ratio $\widehat{\pi}^{-}_{i,q}$, or the equal-performance ratio $\widehat{\pi}^{0}_{i,q}$ of fund i at quarter q.[23] Model (7) is estimated by nonlinear least squares for each of the 39 estimation samples separately.[24]

The regression results are presented in Panel A of Table 4 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 from zero, significantly positive, and significantly negative at the 5% level, respectively.

Fund size, fund age, and the interaction effect between fund size and fund age are the key variables considered for testing *Hypotheses I* and *II*. The distribution of the significance of the estimated coefficients on the rolling sample provides strong statistical evidence that fund size and fund age are indeed important drivers of peer performance. For the outperformance ratio, we find

[23] A potentially more efficient alternative to the equation-by-equation estimation for $\widehat{\pi}^{+}_{i}$, $\widehat{\pi}^{-}_{i}$ and $\widehat{\pi}^{0}_{i}$ 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 in terms of the model for the dependence during the estimation of the ratios.

[24] 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 $G^{-1}(\widehat{\pi}^{\bullet}_{i,q})$ is also not possible because of the non-negligible proportion of funds for which $\widehat{\pi}^{\bullet}_{i,q}$ is exactly one.

that for 79% 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 85% of the samples, so we conclude that fund size leads to a deterioration in the outperformance ratio.

Similarly, 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 provide empirical support for *Hypothesis I*, which predicts that, because of decreasing returns of scale in active investment strategies, larger funds have fewer opportunities to outperform their smaller peers.

For the underperformance ratio, the effects of fund size and fund age are both important and they interact. We find that the coefficient on fund size is significantly positive for 62% of the samples, the coefficient on fund age is positive for 97% of the samples, and the coefficient on the interaction variable between fund size and fund age is significantly negative for 69% 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 outperformance 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 performance 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, which is consistent with *Hypothesis I*. Third, we find that, in agreement with *Hypothesis II*, the increase in the underperformance ratio for larger (respectively older) funds is partly compensated for by the age (respectively 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.

Results in Panel A of Table 4 are obtained for the ratios benchmarking the hedge funds' performance against the peer funds pursuing the same investment style. In Panel B we show the results of the robustness analysis with respect to the choice of peer group. Instead of restricting it to the funds belonging to the same style (Panel A), we define in Panel B the group of peer hedge funds as the set of all hedge funds in the universe. This introduces heterogeneity in the peer performance evaluation and is therefore likely to reduce the power of the peer performance ratios to separate skilled from unskilled fund managers. In fact, as mentioned by Hunter et al. (2014), controlling for funds pursuing similar strategies tends to improve the selection of funds with future outperformance. The peer performance ratios estimated using all hedge funds in the universe are highly correlated with the peer performance ratios estimated using only the funds pursuing the same investment style as the peer category (see Table2). It is thus not a surprise that the determinants of the former (as reported in Panel B of Table 4) are similar to those of the latter (Panel A). Consistent with the view that controlling for investment style leads to more accurate proxies for the hedge fund managers' skill, the number of significant coefficients is slightly lower for the peer performance ratios estimated using the whole universe as peer group. Overall, both regression results in Panels A and B of Table 4 provide strong empirical support for *Hypotheses I* and *II*.

[Insert Table 4 about here]

4.5. Testing Hypotheses IIIa and IIIb: Fund selection based on the outperformance ratio compared with alternative peer performance measures

Is the outperformance ratio better at forecasting the short-run performance of hedge funds than the fund's alpha, relative alpha (Jagannathan et al., 2010), peer alpha (Hunter et al., 2014), distinctiveness (Sun et al., 2012) and selectivity (Amihud and Goyenko, 2013) measures? We here use an extensive out-of-sample portfolio analysis to answer this question. We still use monthly returns for the estimation of the peer performance parameters, but set the prediction horizon to one quarter in order to balance the longer holding periods required in hedge fund investments. The use of this horizon is also consistent with Agarwal and Naik (2000)'s finding that persistence tends to be maximum at the quarterly horizon, compared with half-yearly and yearly investment horizons.

In order to assess the economic gains of selecting funds based on the peer performance ratios compared with alternative (peer) performance measures, we present in the top plot of Figure 3 the wealth evolution of a quarterly rebalanced investment strategy of investing in the top quintile funds (dashed line) and bottom quintile funds (dotted lines) in terms of the outperformance ratio (with equal weights).[25] The out-of- sample evaluation ranges from 2005-Q1 to 2014-Q2 and has thus 38 rebalancing dates. We see that the top quintile funds in terms of the outperformance ratio systematically outperform the 20% bottom funds in terms of the outperformance ratio. The outperformance ratio is thus useful for distinguishing the funds with a higher than average performance from those with a lower than average performance. Moreover, as can be seen in Panel A of Table 5, the higher cumulative performance of the top quintile outperformance ratio based portfolio is achieved at a lower level of volatility.

[Insert Figure 3 and Table 5 about here]

The outperformance ratio is thus a valuable metric for fund selection, but how does this add value compare with investing in the top quintile funds based on the other peer performance measures? To answer this question, we report the relative performance charts in the middle and bottom plots of Figure 3 and present the summary performance statistics in Panel B of Table 5.[26]

First, let us first consider the relative performance charts. We see that for the overall period, the top quintile outperformance ratio portfolio yields the highest total return. While the total return is only slightly higher compared with the top quintile portfolios using the fund's alpha, peer alpha or relative alpha, its total return is more than 40% higher compared with the top quintile portfolios in terms of selectivity and distinctiveness. This result is also clear in the summary performance measures of Table 5, where we report the annualized average return, volatility, Sharpe ratio and alpha (of the nine-factor model) of the quarterly rebalanced portfolios. The out-of-sample performance results indicate that the distinctiveness and selectivity measures fail at identifying the

[25] For the bottom quintile portfolio in terms of the outperformance ratio, when there is a tie in terms of more than 20% of funds with a zero outperformance rate, we additionally rank on the underperformance ratio.

[26] For brevity in presentation, we limit ourselves to the performance measures in Table 2. We also analyzed fund selection based on the t-statistics associated to the fund's alpha, peer alpha and relative alpha, and obtained similar results as when using the fund's alpha, peer alpha and relative alpha, respectively. Also, the use of decile portfolios instead of quintile portfolios led to similar results.

funds with a higher return in the next period. This may be due to the construction of the distinctiveness and the selectivity measures, which only consider the absolute magnitude of deviation of the fund's performance compared with the benchmark return, and not the direction of out- or underperformance.

Let us next investigate in more detail the full-sample statistics measuring the various facets of portfolio performance, and analyze its robustness with respect to the choice of the weighting method. Table 5 shows the performance of the top and bottom portfolios obtained by sorting the funds on the different measures of (peer) performance, and using both equal-weighting as well as value-weighting.[27] Summary measures on the out-of-sample investment performance obtained by sorting using the outperformance ratio are shown in Panel A, while the performance results obtained using the alternative measures are shown in Panel B. Overall, we see that the top quintile portfolio based on the outperformance ratio has the best risk-adjusted performance of all portfolios considered, both in terms of Jensen's alpha as in terms of the Sharpe ratio. The analysis using portfolio sorting thus confirms *Hypothesis IIIa* about the gains in risk-adjusted performance when the outperformance ratio is used to select the top quintile funds.

The alternative performance measures in Panel B have the disadvantage of not correcting for luck. The outperformance ratio is an alternative for them, which in most cases leads to superior out-of-sample performance when used to construct quintile portfolios. According to *Hypothesis IIIb* the best of two worlds can be achieved by using the outperformance ratio as a second pass filter in selecting hedge funds. This suggestion is tested in Panel C of Table 5, where we present the results of investing in the funds obtained by first selecting the top 40% funds in terms of the competing performance measures (not adjusted for luck) and then selecting the top half of them based on the outperformance ratio (which corrects for luck). In all cases, this leads to a significant reduction in the volatility of the out-of-sample returns, compared with the corresponding top

[27] 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.

quintile portfolio, and, as stated in *Hypothesis IIIb*, it almost always increases the risk-adjusted return.

Finally, in Panel D of Table 5, we redo the analysis but for peer performance measures computed with all funds as peers (denoted by $\widehat{\pi}^{*+}$). We see that the risk-adjusted performance is still significant, but less pronounced. This confirms our recommendation to control for fund style in the peer analysis, and is consistent with Hunter et al. (2014), who show that, controlling for funds pursuing similar strategies, tends to improve the selection of funds with future outperformance.

4.6. Testing Hypothesis IV: The outperformance ratio and future fund performance

The bottom line of the portfolio analysis in the previous section is that, consistent with *Hypotheses IIIa* and *IIIb*, we find that the outperformance ratio is effectively able to select the top performing hedge funds. We analyze now this outcome in further detail and use multivariate regression techniques to control for other influences in the analysis of the predictive power of the outperformance ratio for forecasting fund performance one quarter in the future.

The dependent variable $R_{i,q}$ is the quarterly net return of hedge fund i at the end of quarter q. To save space, we focus on the comparison between the outperformance ratio ($\widehat{\pi}^{+}_{i,q-1}$) and the peer alpha ($\widehat{\alpha}^{\text{peer}}_{i,q-1}$) and relative alpha ($\widehat{\alpha}^{\text{rel}}_{i,q-1}$) as predictor variables for future returns.[28] Consistent with *Hypothesis IIIb* about the complementarity of the outperformance ratio and the relative (or peer) alpha, we also include the interaction variable between these two measures (*i.e.*, $IT^{\text{rel/peer}}_{i,q-1} \equiv \widehat{\alpha}^{\text{rel/peer}}_{i,q-1} \times \widehat{\pi}^{+}_{i,q-1}$).

In Table 6 we report the least squares estimation results for various regressions, which are

[28] As could be expected from the portfolio analysis results, the distinctiveness and selectivity measures are also in the regression analysis not a predictor of larger than expected next quarter returns. To simplify the exposition, and for sake of model parsimony, we therefore do not include them in the regression analysis.

nested in the generalized unrestricted model given by:

$$R_{i,q} = \zeta_q + \beta_1 \widehat{\pi}^{+}_{i,q-1} + \beta_2 \widehat{\alpha}^{\text{rel}}_{i,q-1} + \beta_3 \widehat{\alpha}^{\text{peer}}_{i,q-1} + \beta_4 \boldsymbol{IT}^{\text{rel}}_{i,q-1} + \beta_5 \boldsymbol{IT}^{\text{peer}}_{i,q-1} + \boldsymbol{\gamma}' \boldsymbol{CTRL}_{i,q-1} + \varepsilon_{i,q} \,, \quad (8)$$

where ζ_q is the quarter's fixed effect, $\varepsilon_{i,q}$ is the error term and $\boldsymbol{\gamma}$ is the (13×1) vector of parameters on the control variables:

$$\boldsymbol{CTRL}_{i,q-1} \equiv \big(R_{i,q-1}, LAUM_{i,q-1}, AGE_{i,q-1}, F_{i,q-1}, \\ VOL_{i,q-1}, EH_i, MA_i, RV_i, MF_i, PF_i, LEV_i, HUR_i, HWM_i\big)' \,.$$

All control variables are lagged to avoid look-ahead bias in our predictive regression model and there is a total of 46,378 fund-quarter observations in the estimation sample.[29]

The main results can be summarized in two important findings. First, for all specifications considered, the outperformance ratio has a positive and significant predictive power for the one-quarter ahead return. Second, the forecast performance is improved when including an interaction term with either relative alpha or peer alpha. In both cases, the coefficient on the interaction effect is positive, confirming the complementarity of the outperformance ratio and the peer (or relative) alpha. The higher the relative alpha, the larger is the increase in the predicted return when the outperformance ratio increases. The multivariate regression result thus confirms the conclusions of the portfolio analysis and *Hypothesis IV*: The outperformance ratio is a good predictor of future fund performance and this result is robust to the different control variables. The effects on the control variables are as expected. There is a positive and significant effect of past returns on future

[29] As control variables, we include the fund's lagged fund return to account for the positive autocorrelation in quarterly hedge fund returns (see Agarwal and Naik, 2000); to control for fund size effects, we include the assets under management (in logarithm), together with the fund's capital inflow defined by Fung et al. (2008) as $F_{i,q-1} \equiv (AUM_{i,q-1} - AUM_{i,q-2}(1 + R_{i,q-1}))/(AUM_{i,q-2})$. 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 as used in (7), namely the investment style, the use of leverage, the existence of a hurdle rate and other time-invariant variables in fee structure.

returns, older and larger funds tend to have a smaller return, and funds with a higher volatility tend to offer a higher return.

[Insert Table 6 about here]

5. Conclusion

The peer performance of active fund managers is a topic of importance for both academics and practitioners. Existing approaches to compare a fund's performance with the performance of its peers do not adjust for the possibility that the difference in performance may be due to chance. 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 develop an estimator for the corresponding equal-, out-, and underperformance ratios.

We then 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 demonstrate that equal-performance dominates out- and underperformance of hedge funds with respect to their peers. As a consequence, on average, percentile-rank analyses of out- and underperformance are too optimistic about the outperformance of the funds with a relatively good ranking and too pessimistic about the underperformance of the funds with a worse ranking.

We show that the cross-sectional variation in the peer performance ratios is to a large extent determined by the fund size. As predicted by the hedge fund life-cycle theory of Berk and Green (2004) and Getmansky (2012), we find that fund size has a negative impact on the outperformance ratio. It also tends to increase the underperformance ratio, but to a larger extent for young funds than for old funds. The interaction effect between the fund's size and age is consistent with the career hypothesis that large funds tend to herd more and that the propensity to herd increases with the fund's age (see, *e.g.*, Scharfstein and Stein, 1990; Brown et al., 2001).

Finally, in the out-of-sample evaluation using rolling five-year estimation samples, we use portfolio sorts and multivariate regressions to analyze the usefulness of the outperformance ratio as a predictive variable to identify the next quarter's outperforming funds. The investment strategy of investing in the quarterly rebalanced equally-weighted (and value-weighted) portfolio of the top quintile outperformance ratio hedge funds has a substantially higher total and risk-adjusted performance than the top quintile portfolio based on the fund's distinctiveness and selectivity measures proposed by Sun et al. (2012) and Amihud and Goyenko (2013). Importantly, the best portfolio risk-adjusted performance is achieved by the joint use of the outperformance ratio with either the relative alpha of Jagannathan et al. (2010) or the peer alpha of Hunter et al. (2014). This positive interaction effect of the outperformance ratio and the peer or relative alpha is confirmed in a multivariate regression analysis, where we find that the outperformance ratio has a significant positive predictive value for the next quarter's hedge fund return and that this effect increases when the alpha or relative alpha values are higher. This result is robust to the inclusion of various control variables such as the fund's lagged fund return and volatility, the fund size and age.

References

Agarwal, V., Naik, N., 2000. Multi-period performance persistence analysis of hedge funds. Journal of Financial and Quantitative Analysis 35, 327–342. doi:10.2307/2676207.

Aggarwal, R.K., Jorion, P., 2010. The performance of emerging hedge funds and manager. Journal of Financial Economics 96, 238–256. doi:10.1016/j.jfineco.2009.12.010.

Amihud, Y., Goyenko, R., 2013. Mutual fund's R2 as predictor of performance. Review of Financial Studies 26, 667–694. doi:10.1093/rfs/hhs182.

Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858. doi:10.2307/2938229.

Andrews, D., Monahan, J., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 60, 953–966. doi:10.2307/2951574.

Aragon, G., Liang, B., Park, H., 2013. Onshore and offshore hedge funds: are they twins? Management Science 60, 74–91. doi:10.1287/mnsc.2013.1729.

Ardia, D., Boudt, K., 2015. Testing equality of modified Sharpe ratios. Finance Research Letters 13, 97–104. doi:10.1016/j.frl.2015.02.008.

Ardia, D., Boudt, K., 2016. **PeerPerformance**: Peer performance analysis in R. URL: https://github.com/ArdiaD/PeerPerformance. version 2.0.8.

Bajgrowicz, P., Scaillet, O., 2012. Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics 106, 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. Journal of Finance 65, 179–216. doi:10.1111/j.1540-6261.2009.01527.x.

Berk, J., Van Binsbergen, J.H., 2015. Measuring skill in the mutual fund industry. Journal of Financial Economics 118, 1–20. doi:10.1016/j.jfineco.2015.05.002.

Berk, J.B., Green, R.C., 2004. Mutual fund flows and performance in rational markets. Journal of Political Economy 112, 1269–1295. doi:10.1086/424739.

Boyson, N.M., 2010. Implicit incentives and reputational herding by hedge fund managers. Journal of Empirical Finance 17, 283–299. doi:10.1016/j.jempfin.2009.10.005.

Brown, S.J., Goetzmann, W.N., 1997. Mutual fund styles. Journal of Financial Economics 43, 373–399. doi:10.1016/S0304-405X(96)00898-7.

Brown, S.J., Goetzmann, W.N., 2003. Hedge funds with style. The Journal of Portfolio Management 29, 101–112.

Brown, S.J., Goetzmann, W.N., Park, J., 2001. Careers and survival: Competition and risk in the hedge fund and CTA industry. Journal of Finance 56, 1869–1886. doi:10.1111/0022-1082.00392.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 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. American Economic Review 94, 1276–1302. doi:10.1257/0002828043052277.

Chen, Y., Cliff, M.T., Zhao, H., 2015. Hedge funds: The good, the bad, and the lucky. Journal of Financial and Quantitative Analysis, forthcoming .

Cremers, K.M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. Review of Financial Studies 22, 3329–3365. doi:10.1093/rfs/hhp057.

Fama, E., French, K., 2010. Luck versus skill in the cross-section of mutual fund returns. Journal of Finance 65, 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. URL: http://www-bcf.usc.edu/~ferson/papers/yong.pdf.

Fung, W., Hsieh, D., 2004. Hedge fund benchmarks: A risk based approach. Financial Analysts Journal 60, 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. Journal of Finance 63, 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. Quarterly Journal of Finance 2, 1–53. doi:10.1142/S2010139212500036.

Graham, J., 1999. Herding among investment letters: Theory and evidence. Journal of Finance 54, 237–268. doi:10.1111/0022-1082.00103.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365–380.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497. doi:10.3982/ECTA5771.

HFR, 2016. Hedge fund assets rise on Brexit gains. Hedge Fund Research Inc. Press Release July, 1–3. URL: https://www.hedgefundresearch.com/sites/default/files/pdf/pr_20150120.pdf.

Hunter, D., Kandel, E., Kandel, S., Wermers, R., 2014. Mutual fund performance evaluation with active peer benchmarks. Journal of Financial Economics 112, 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. Journal of Finance 65, 217–255. doi:10.1111/j.1540-6261.2009.01528.x.

Joenväärä, J., Kosowsk, R., Tolonen, P., 2016. Hedge fund performance: What do we know? Working paper. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1989410.

Kosowski, R., Naik, N., Teo, M., 2005. Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. Journal of Financial Economics 84, 229–264. doi:10.1016/j.jfineco.2005.12.009.

Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the Sharpe ratio. Journal of Empirical Finance 15, 850–859. doi:10.1016/j.jempfin.2008.03.002.

Liang, B., 1999. On the performance of hedge funds. Financial Analysts Journal 55, 72–85. doi:10.2469/faj.v55.n4.2287.

Pástor, L., Stambaugh, R.F., 2002. Mutual fund performance and seemingly unrelated assets. Journal of Financial Economics 63, 315–349. doi:10.1016/S0304-405X(02)00064-8.

Pástor, L., Stambaugh, R.F., Taylor, L.A., 2015. Scale and skill in active management. Journal of Financial Economics 116, 23–45. doi:10.1016/j.jfineco.2014.11.008.

Pav, S.E., 2014. **SharpeR**: Statistical significance of Sharpe ratio in R. URL: https://github.com/shabbychef/SharpeR. version 0.1401.

R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Rosenblatt, M., 1952. Remarks on a multivariate transformation. Annals of Mathematical Statistics 23, 470–472. doi:10.1214/aoms/1177729394.

Scharfstein, D.S., Stein, J.C., 1990. Herd behavior and investment. American Economic Review 80, 465–479.

Storey, J., 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society B 64, 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. Review of Financial Studies 25, 96–143. doi:10.1093/rfs/hhr092.

Titman, S., Tiu, C., 2011. Do the best hedge funds hedge? Review of Financial Studies 24, 123–168. doi:10.1093/rfs/hhq105.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126. doi:10.1111/1468-0262.00152.

Table 1: Summary statistics

This table presents the average summary statistics for the universe composition, fund performance, and fund characteristics over the quarterly updated five-year estimation samples between January 2000 and June 2014 (39 samples). Panel A reports the annualized performance (geometric return in percent) of 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 fund monthly alpha values (in percent, annualized). Panel C reports the distribution 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. Management and performance fees are in percent. Leverage, hurdle rate, and high watermark are dummy variables. See Section 4.1 for further details.

	N	Mean	Med	Std	Min	25th	75th	Max
Panel A: Fund compound returns (in percent, annualized)								
Equity Hedge	715	8.53	7.53	9.65	-32.90	3.48	12.48	71.68
Event-Driven	163	8.65	7.83	7.85	-16.49	4.35	11.86	49.24
Macro	290	8.13	7.04	8.41	-19.28	3.25	11.59	58.15
Relative Value	224	8.67	7.92	8.71	-25.84	4.60	11.90	57.24
Panel B: Fund monthly alpha (in percent, annualized)								
Equity Hedge		3.08	2.54	8.41	-37.59	-1.23	6.89	57.15
Event-Driven		3.94	3.49	6.82	-18.34	0.42	7.01	37.68
Macro		5.86	4.88	9.23	-24.20	0.57	9.90	62.07
Relative Value		5.03	4.41	7.80	-26.66	1.33	8.28	47.32
Panel C: Fund characteristics								
AUM (mn USD)		200.29	45.18	538.22	0.02	11.92	158.54	8939.37
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 (y/n)		0.63	1.00					
Hurdle rate (y/n)		0.13	0.00					
High watermark (y/n)		0.91	1.00					

Table 2: Correlation matrix

This table presents the correlation matrix between the various peer performance measures: the estimated alpha of a nine-factor model ($\widehat{\alpha}$), the estimated relative alpha ($\widehat{\alpha}^{\text{rel}}$), the estimated peer alpha ($\widehat{\alpha}^{\text{peer}}$), the estimated distinctiveness ($\widehat{D}$), the estimated selectivity ($\widehat{S}$), the outperformance ratio ($\widehat{\pi}^{+}$) and the outperformance ratio computed with all funds as the peer universe ($\widehat{\pi}^{*+}$). The upper triangular part reports Pearson and the lower triangular part reports Spearman correlation coefficients (in percent) between the 46,378 estimates of the various peer performance measures, obtained over the 39 estimation samples in our January 2000–June 2014 HFR database of dead and alive Equity Hedge, Event-Driven, Macro, and Relative Value hedge funds. See Section 4.1 for further details.

	$\widehat{\alpha}$	$\widehat{\alpha}^{\text{rel}}$	$\widehat{\alpha}^{\text{peer}}$	$\widehat{D}$	$\widehat{S}$	$\widehat{\pi}^{+}$	$\widehat{\pi}^{*+}$
$\widehat{\alpha}$		43.4	74.5	11.9	25.0	76.7	79.1
$\widehat{\alpha}^{\text{rel}}$	62.2		61.5	20.6	13.2	34.3	33.1
$\widehat{\alpha}^{\text{peer}}$	70.9	87.8		34.4	22.2	58.5	56.9
$\widehat{D}$	17.2	37.9	41.8		53.7	8.6	9.9
$\widehat{S}$	28.5	24.8	27.3	68.3		20.2	23.0
$\widehat{\pi}^{+}$	75.6	55.7	63.1	11.8	19.8		96.0
$\widehat{\pi}^{*+}$	77.4	54.7	62.1	14.6	22.5	87.1	

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 4.2 for further 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[0.67;0.00]	1[0.09;0.00]	2[0.46;0.00]	3[0.00;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]

Table 4: Estimation results of the nonlinear regressions

This table presents the summary of the estimation results for the nonlinear model (7) with the dependent variable $\widehat{\pi}^{\bullet}$ (computed over the 39 samples). Panel A reports results for peer ratios computed using funds with the same strategy as peers (*i.e.*, $\widehat{\pi}^{+}$, $\widehat{\pi}^{-}$, $\widehat{\pi}^{0}$). Panel B reports results for peer ratios computed using all funds in the universe as peers (*i.e.*, $\widehat{\pi}^{*+}$, $\widehat{\pi}^{*-}$, $\widehat{\pi}^{*0}$). For each peer performance ratio, the table shows the average value $\overline{\beta}$ (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, $LAUM_{q-1}$), the fund's age in months (AGE_{q-1}), and their interaction ($IT_{q-1} \equiv LAUM_{q-1} \times AGE_{q-1}$), at the end of the preceding sample. *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. 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 4.4 for further details.

Panel A: Peer ratios computed with same strategy funds as peers												
	$\widehat{\pi}^{+}$				$\widehat{\pi}^{-}$				$\widehat{\pi}^{0}$			
	$\overline{\beta}$	$\neq 0$	> 0	< 0	$\overline{\beta}$	$\neq 0$	> 0	< 0	$\overline{\beta}$	$\neq 0$	> 0	< 0
Constant	−1.93	100	0	100	−0.77	79	0	82	0.12	21	23	15
$LAUM_{q-1}$	−0.10	72	0	79	0.08	46	62	5	0.04	36	49	0
AGE_{q-1}	−0.04	28	0	36	0.07	92	97	0	−0.01	8	3	13
IT_{q-1}	0.00	15	15	0	−0.01	46	0	69	0.00	5	5	3
EH	−0.25	56	5	62	−0.44	85	0	85	0.48	82	85	0
MA	−0.65	54	0	59	−0.54	56	0	64	0.65	72	69	8
RV	0.08	31	23	18	−0.13	23	10	38	0.01	54	31	26
MF	38.37	79	87	0	−31.77	69	0	69	−3.71	13	3	28
PF	1.67	49	46	10	−1.69	41	0	51	0.03	8	8	8
LEV	−0.07	33	10	33	−0.01	5	0	10	0.05	23	21	3
HUR	0.11	33	31	10	0.11	3	13	0	−0.11	31	3	31
HWM	0.32	26	31	0	−0.21	26	0	38	0.02	0	0	0
Pseudo R^2		96.65				97.11				96.33		

Panel B: Peer ratios computed with all funds as peers												
	$\widehat{\pi}^{*+}$				$\widehat{\pi}^{*-}$				$\widehat{\pi}^{*0}$			
	$\overline{\beta}$	$\neq 0$	> 0	< 0	$\overline{\beta}$	$\neq 0$	> 0	< 0	$\overline{\beta}$	$\neq 0$	> 0	< 0
Constant	−2.10	100	0	100	−1.23	87	0	90	0.55	64	77	0
$LAUM_{q-1}$	−0.10	74	0	79	0.08	54	51	5	0.05	38	54	0
AGE_{q-1}	−0.03	18	3	23	0.08	87	97	0	−0.02	10	0	28
IT_{q-1}	0.00	13	13	8	−0.01	36	0	49	0.00	3	5	0
EH	−0.21	64	13	54	0.17	54	44	18	0.08	26	28	5
MA	0.01	90	44	49	−0.41	69	18	56	0.13	62	49	21
RV	0.34	56	59	3	−0.34	38	3	59	−0.06	23	3	26
MF	31.94	77	87	0	−33.43	69	0	74	−3.06	10	5	15
PF	1.57	54	51	10	−1.38	31	0	31	−0.19	15	10	13
LEV	−0.08	38	15	38	−0.04	18	3	15	0.07	31	33	5
HUR	0.10	31	28	5	0.01	3	8	0	−0.08	23	3	31
HWM	0.30	21	33	0	−0.18	23	0	36	0.00	0	0	3
Pseudo R^2		95.49				96.60				98.92		

Table 5: Performance results of the quintile portfolios

This table presents the annualized return (Mean, in percent), volatility (Std, in percent), Sharpe ratio and 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. Panel A reports results for portfolios invested in the bottom and top quintiles of the hedge funds ranked on their past outperformance ratio computed using funds with the same strategy as peers ($\widehat{\pi}^{+}$). Panel B reports results for portfolios invested in the bottom and top quintile of the hedge funds ranked on their alpha ($\widehat{\alpha}$), relative alpha ($\widehat{\alpha}^{\text{rel}}$), peer alpha ($\widehat{\alpha}^{\text{peer}}$), distinctiveness ($\widehat{D}$) and selectivity ($\widehat{S}$). All measures are computed on five-year rolling samples of monthly net returns. Panel C reports results for portfolios invested in the hedge funds that belong to the top half in term of the outperformance ratio $\widehat{\pi}^{+}$ of the 40% funds with highest alternative performance measure. Those are thus top quintile hedge funds portfolios obtained by sequential ranking. Panel D reports results for portfolios invested in the bottom and top quintile of the hedge funds ranked on their past outperformance ratio computed using all funds in the universe as peers ($\widehat{\pi}^{*+}$). 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 4.5 for further details.

		Equally-weighted portfolios				Value-weighted portfolios			
Criterion	Quintile	Mean	Std	Sharpe	Alpha	Mean	Std	Sharpe	Alpha
Panel A: Portfolio sorts using the outperformance ratio computed with same strategy funds as peers									
$\widehat{\pi}^{+}$	Bottom	5.30	11.63	0.46	0.16	3.22	9.20	0.35	-0.03
	Top	9.88	8.47	1.17	6.00^{***}	7.29	7.97	0.91	4.10^{**}
Panel B: Portfolio sorts using alternative indicators									
$\widehat{\alpha}$	Bottom	6.57	12.96	0.51	0.25	5.03	11.06	0.45	0.75
	Top	9.51	9.13	1.04	5.25^{**}	6.43	8.47	0.76	2.72
$\widehat{\alpha}^{\text{rel}}$	Bottom	6.68	12.67	0.53	0.57	6.09	10.52	0.58	1.70
	Top	8.95	8.66	1.03	5.25^{***}	6.91	8.45	0.82	4.08^{**}
$\widehat{\alpha}^{\text{peer}}$	Bottom	6.89	13.13	0.52	0.35	6.45	10.93	0.59	1.02
	Top	9.03	7.60	1.19	5.76^{***}	6.87	7.46	0.92	4.41^{**}
$\widehat{D}$	Bottom	9.00	14.34	0.63	2.29	8.68	11.53	0.75	3.65^{**}
	Top	4.53	2.47	1.84	3.85^{***}	2.45	3.53	0.69	1.29^{**}
$\widehat{S}$	Bottom	8.62	14.11	0.61	1.42	8.74	12.37	0.71	2.78^{**}
	Top	5.42	4.08	1.33	3.86^{***}	3.53	4.89	0.72	1.99^{*}
Panel C: Portfolio sorts using the outperformance ratio, after sorting on the alternative indicators									
$\widehat{\pi}^{+} \mid \widehat{\alpha}$	Top	9.89	8.52	1.16	5.99^{***}	7.28	7.98	0.91	4.09^{**}
$\widehat{\pi}^{+} \mid \widehat{\alpha}^{\text{rel}}$	Top	9.54	7.85	1.22	5.79^{***}	6.57	7.78	0.84	3.45^{**}
$\widehat{\pi}^{+} \mid \widehat{\alpha}^{\text{peer}}$	Top	9.71	8.02	1.21	6.09^{***}	6.88	7.57	0.91	3.94^{**}
$\widehat{\pi}^{+} \mid \widehat{D}$	Top	6.84	3.89	1.76	5.21^{***}	5.09	5.53	0.92	3.50^{**}
$\widehat{\pi}^{+} \mid \widehat{S}$	Top	6.56	5.47	1.20	4.49^{***}	4.97	6.13	0.81	2.82^{**}
Panel D: Portfolio sorts using the outperformance ratio computed with all funds as peers									
$\widehat{\pi}^{*+}$	Bottom	6.15	11.84	0.52	0.60	4.78	9.37	0.51	1.18
	Top	9.25	8.21	1.13	5.43^{***}	6.33	7.69	0.82	3.13^{*}

Table 6: Estimation results of the multivariate regressions

This table presents the regression results of the return prediction model (8). The dependent variable is the quarterly net return of hedge fund i at the end of quarter q. The explanatory variables of interest are the lagged value of: (i) the fund's outperformance ratio (*i.e.*, $\widehat{\pi}^{+}_{i,q-1}$) computed using the funds pursuing the same style, (ii) the relative alpha (*i.e.*, $\widehat{\alpha}^{\text{rel}}_{i,q-1}$), (iii) the peer alpha (*i.e.*, $\widehat{\alpha}^{\text{peer}}_{i,q-1}$), (iv) the interaction between the relative alpha and the outperformance ratio (*i.e.*, $IT^{\text{rel}}_{i,q-1} \equiv \widehat{\alpha}^{\text{rel}}_{i,q-1} \times \widehat{\pi}^{+}_{i,q-1}$), and (v) the interaction between the peer alpha and the outperformance ratio (*i.e.*, $IT^{\text{peer}}_{i,q-1} \equiv \widehat{\alpha}^{\text{peer}}_{i,q-1} \times \widehat{\pi}^{+}_{i,q-1}$). The control variables are the lagged value of the fund return $R_{i,q-1}$, the fund's asset under management (in logarithm, $LAUM_{i,q-1}$), the fund's age $AGE_{i,q-1}$, the fund's capital inflow $F_{i,q-1}$, the fund's return volatility (computed as the standard deviation over the last twelve months, $VOL_{i,q-1}$), as well as the fund's characteristics described in Table 4 (see page 42). Robust standard errors are shown in parentheses. (Adjusted) R^2 are the (adjusted) R-squared (in percent) evaluating the explanatory power of all variables. The symbols ***, ** and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Overall, 46,378 observations are used in the estimation of the regressions. See Section 4.6 for further details.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\widehat{\pi}^{+}_{i,q-1}$		0.0075^{***}	0.0071^{***}	0.0056^{***}	0.0056^{***}	0.0061^{***}	0.0061^{***}
		(0.0015)	(0.0016)	(0.0018)	(0.0018)	(0.0017)	(0.0017)
$\widehat{\alpha}^{\text{rel}}_{i,q-1}$			0.0327		0	−0.0096	
			(0.0388)		(0.0464)	(0.0456)	
$\widehat{\alpha}^{\text{peer}}_{i,q-1}$				0.1259^{*}	0.1259		−0.0096
				(0.0694)	(0.0834)		(0.0456)
$IT^{\text{rel}}_{i,q-1}$						0.0021^{*}	
						(0.0011)	
$IT^{\text{peer}}_{i,q-1}$							0.0021^{*}
							(0.0011)
$R_{i,q-1}$	0.1102^{***}	0.1073^{***}	0.1069^{***}	0.1062^{***}	0.1062^{***}	0.1071^{***}	0.1071^{***}
	(0.0047)	(0.0048)	(0.0048)	(0.0048)	(0.0048)	(0.0048)	(0.0048)
$LAUM_{i,q-1}$	0.0352^{*}	0.0166	0.0155	0.0132	0.0132	0.0181	0.0181
	(0.0201)	(0.0202)	(0.0202)	(0.0201)	(0.0201)	(0.0203)	(0.0203)
$AGE_{i,q-1}$	−0.0128	−0.0081	−0.0077	−0.0065	−0.0065	−0.0074	−0.0074
	(0.0082)	(0.0082)	(0.0081)	(0.0081)	(0.0081)	(0.0082)	(0.0082)
$F_{i,q-1}$	0	−0.0001	−0.0001	−0.0001	−0.0001	−0.0001	−0.0001
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
$VOL_{i,q-1}$	0.3574^{***}	0.3534^{***}	0.3551^{***}	0.3602^{***}	0.3602^{***}	0.3529^{***}	0.3529^{***}
	(0.0138)	(0.0137)	(0.0139)	(0.0143)	(0.0143)	(0.014)	(0.014)
EH_i	0.4162^{***}	0.3947^{***}	0.3988^{***}	0.4128^{***}	0.4128^{***}	0.4001^{***}	0.4001^{***}
	(0.1191)	(0.1169)	(0.1168)	(0.1165)	(0.1165)	(0.1169)	(0.1169)
MA_i	-0.6925^{***}	-0.6838^{***}	-0.6826^{***}	-0.6861^{***}	-0.6861^{***}	-0.6758^{***}	-0.6758^{***}
	(0.0986)	(0.097)	(0.0968)	(0.0963)	(0.0963)	(0.097)	(0.097)
RV_i	0.4014^{***}	0.3709^{***}	0.3773^{***}	0.3947^{***}	0.3947^{***}	0.3721^{***}	0.3721^{***}
	(0.1088)	(0.1072)	(0.1072)	(0.1073)	(0.1073)	(0.1074)	(0.1074)
MF_i	6.871	3.902	3.864	4.115	4.115	3.695	3.695
	(6.746)	(6.637)	(6.621)	(6.585)	(6.587)	(6.629)	(6.629)
PF_i	-2.366^{***}	-2.568^{***}	-2.557^{***}	-2.553^{***}	-2.553^{***}	-2.575^{***}	-2.575^{***}
	(0.8116)	(0.798)	(0.7964)	(0.7924)	(0.7926)	(0.7971)	(0.7971)
LEV_i	−0.0716	−0.0655	−0.0656	−0.0633	−0.0633	−0.0648	−0.0648
	(0.0785)	(0.0771)	(0.0769)	(0.0765)	(0.0765)	(0.077)	(0.077)
HUR_i	0.1506	0.1384	0.1355	0.134	0.134	0.1304	0.1304
	(0.1092)	(0.1071)	(0.1069)	(0.1063)	(0.1063)	(0.107)	(0.107)
HWM_i	0.1438	0.1311	0.13	0.129	0.129	0.1298	0.1298
	(0.1483)	(0.1454)	(0.1451)	(0.1443)	(0.1443)	(0.1452)	(0.1452)
R^2	2.86	2.90	2.90	2.91	2.91	2.91	2.91
Adjusted R^2	2.83	2.87	2.87	2.87	2.87	2.87	2.87

Figure 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 4.2 for further details.

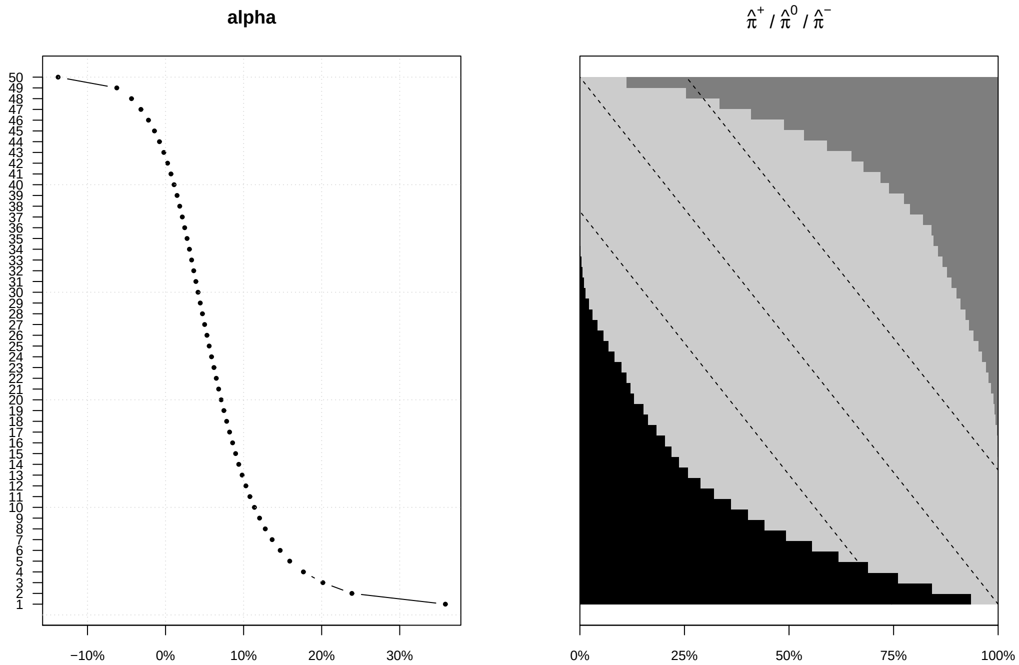


Figure 2: Luck-adjustment correction

The outperformance (resp. underperformance) ratio is a luck-adjusted alternative to the common practice of measuring outperformance (resp. underperformance) as the percentage of peer funds for which the fund's performance is lower (resp. higher) than the fund of interest. This figure presents the average values of the luck-adjustment correction terms for the hedge funds in our database ranked by decreasing alpha and grouped in 50 buckets. The left plot displays the correction for luck in percentile-rank estimates of outperformance ratio. The right plot displays the correction for luck in percentile-rank estimates of the underperformance ratio. Peer performance ratios used to compute $\widehat{\delta}^{+}$ and $\widehat{\delta}^{-}$ 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 4.3 for further details.

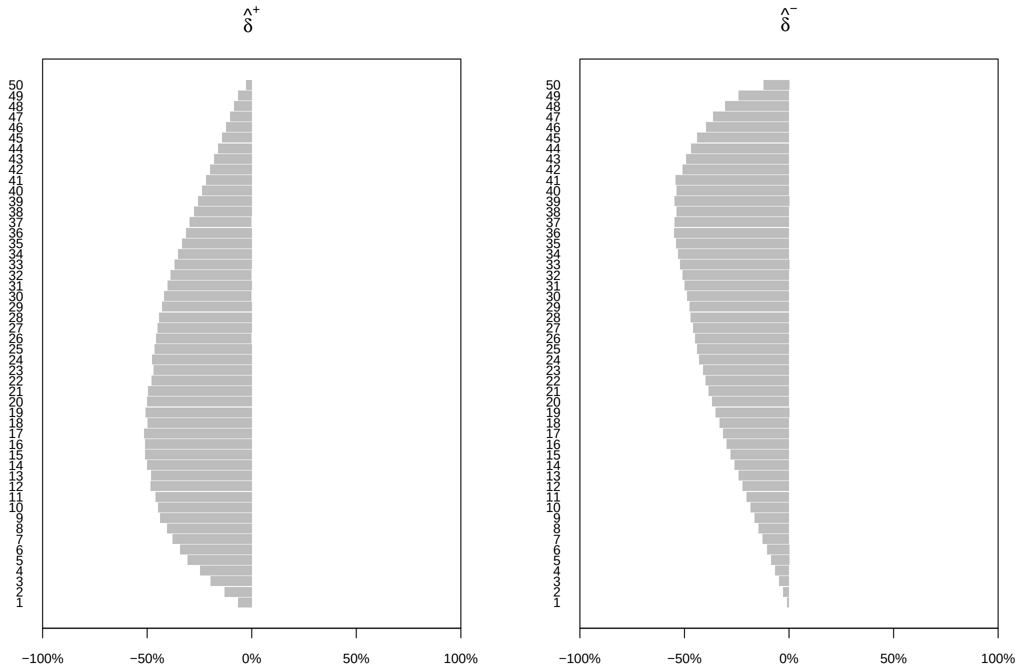
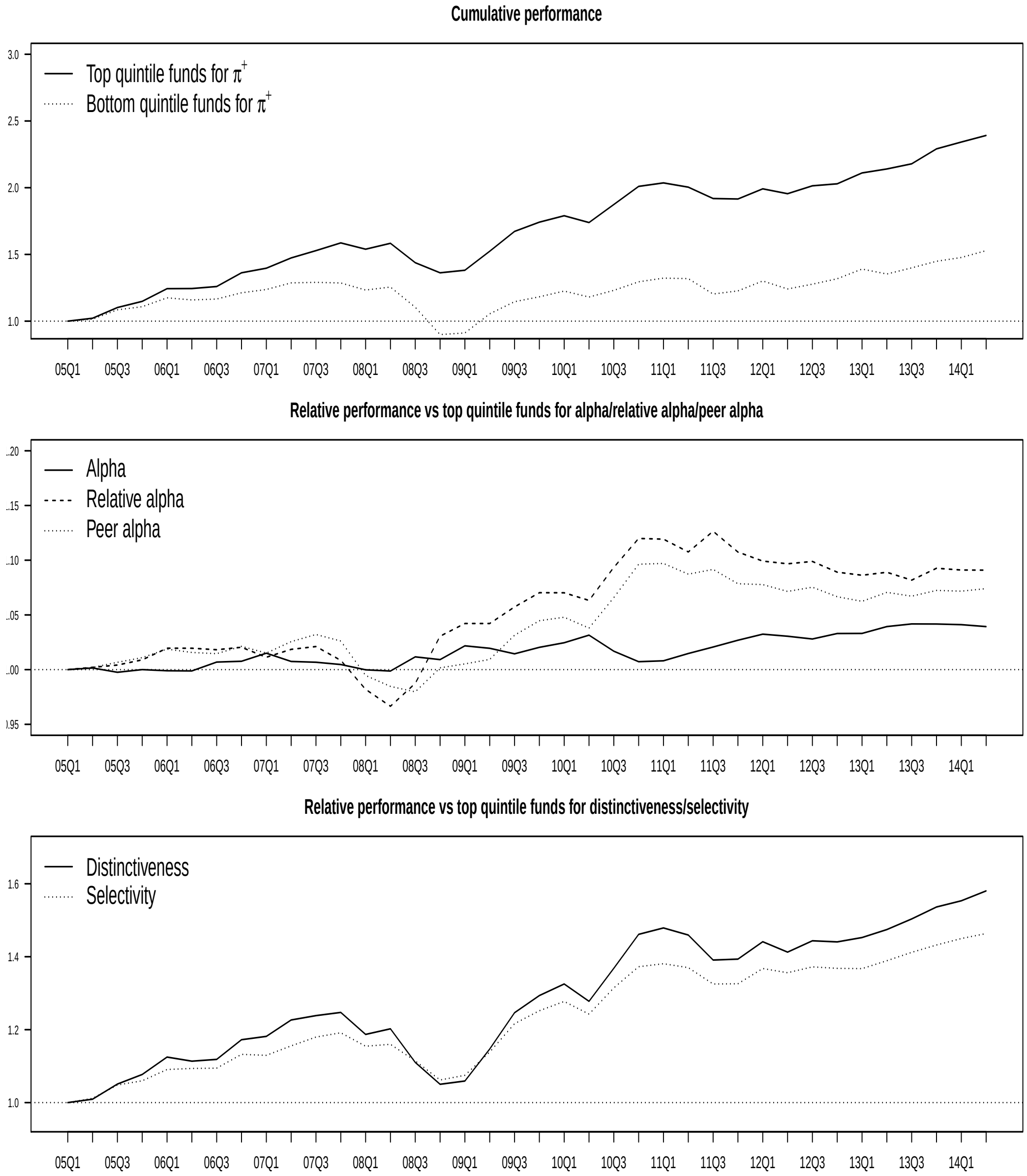


Figure 3: Performance results of the quintile portfolios

This figure presents the cumulative (and relative) performance of quintile portfolios. The top plot displays the cumulative performance of portfolios invested in top (solid line) and bottom (dotted line) quintile funds in terms of the estimated (past) outperformance ratio $\widehat{\pi}^{+}$. The middle plot displays the relative cumulative performance of investing in the top quintile funds according to $\widehat{\pi}^{+}$ with respect to the investment in the top quintile funds based on alpha (solid line), relative alpha (dashed line) and peer alpha (dotted line). The bottom plot displays the relative cumulative performance of investing in the top quintile funds according to $\widehat{\pi}^{+}$ with respect to the investment in the top quintile funds based on distinctiveness (solid line) and selectivity (dotted line). The out-of-sample evaluation period ranges from 2005-Q1 to 2014-Q2 for a total of 38 rebalancing dates. See Section 4.5 for further details.



Appendix A. Bias correction factor

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

$$\widetilde{\pi}_i^0 \equiv \min\left\{\widehat{p}_i^0, 1\right\} ,$$

with:

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

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

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

with:

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

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

Since $\widetilde{\pi}_i^0$ does not estimate π_i^0 but $h(\pi_i^0)$, we use $h^{-1}(\widetilde{\pi}_i^0)$ to estimate π_i^0. Equivalently, we define the correction factor c_i^0 in (4) as $c_i^0 \equiv h^{-1}(\widetilde{\pi}_i^0)/\widetilde{\pi}_i^0$.

Intuitively, the results in (A.1)–(A.2) are obtained because the randomness in $\widetilde{\pi}_i^0$ 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 (2), all p-values exceeding λ_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 λ_i. Then, the actual number of p-values exceeding λ_i, *i.e.*, $\sum_{j \neq i} I\{\widehat{p}_{i-j} \geq \lambda_i\}$, 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:

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

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

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

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

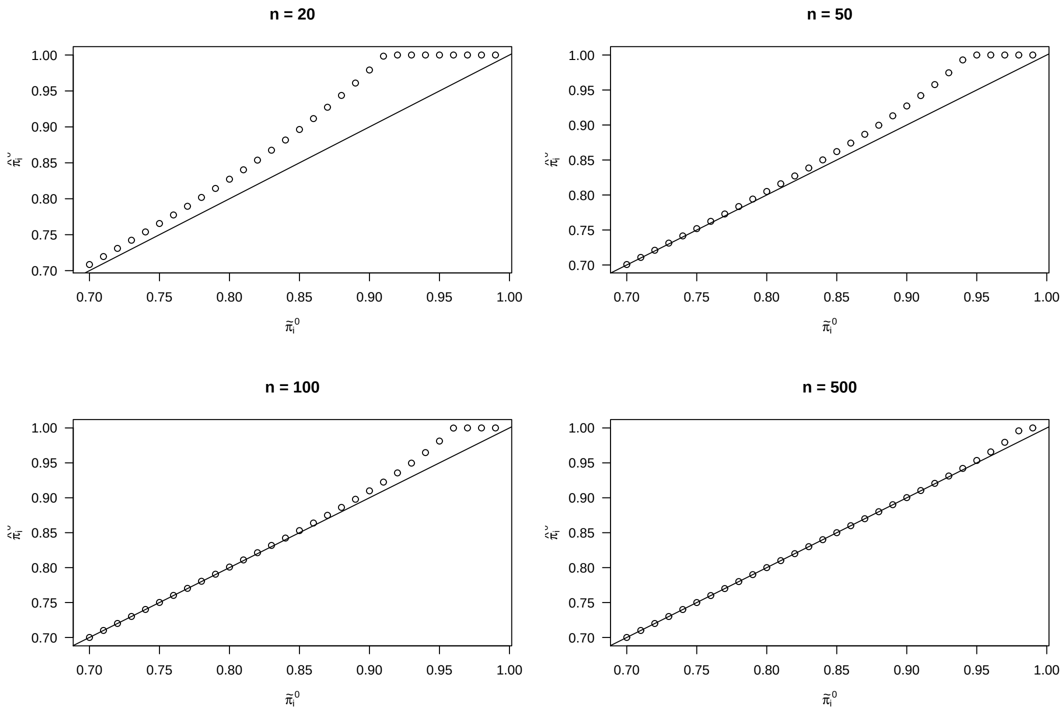
$$\mathbb{E}[\min\{Z,k\}] = \int_{-\infty}^{k} z\phi(z)dz + k(1-\Phi(k))$$
$$= -\int_{-\infty}^{k} \phi'(z)dz + k(1-\Phi(k)) = -\phi(k) + k(1-\Phi(k)),$$

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

Since the non-truncated equal-performance ratio $\widehat{p}_i^0$ is asymptotically normally distributed around the true equal-performance ratio π_i^0, the probability that it exceeds one (and hence the extent of correction needed) increases when π_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 Figure A.4 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 $\widetilde{\pi}_i^0$ is large or n is small.

Figure A.4: Impact of the correction factor

This figure presents the plot of $\widetilde{\pi}_i^0$ (horizontal axis; without correction factor) versus $\widehat{\pi}_i^0$ (vertical axis; with correction factor) for various sizes n of the peer universe.



Appendix B. Choice of threshold value $\boldsymbol{\lambda}_i$

Based on Storey (2002), Barras et al. (2010, footnote 10) proposed a bootstrap procedure for determining the value of λ_i in a purely data-driven manner, which minimizes the estimated mean squared error (MSE) of $\widehat{\pi}_i^0$. Their approach can also be applied to our proposed peer performance analysis. More specifically, we choose λ_i such that an estimate of the MSE of $\widehat{\pi}_i^0(\lambda)$ defined as $\mathbb{E}[(\widehat{\pi}_i^0(\lambda) - \pi_i^0)^2]$ is minimized. First, we compute $\widehat{\pi}_i^0(\lambda)$ using (5) across a range of λ values ($\lambda \in \{0.3, 0.32, \ldots, 0.7\}$). Second, for each possible value of λ, we form B bootstrap replicates of $\widehat{\pi}_i^0(\lambda)$ by drawing with replacement from the $(n \times 1)$ vector of fund p-values. These are denoted by $\widehat{\pi}_i^{0*b}(\lambda)$ for $b = 1, \ldots, B$. Third, we compute the estimated MSE for each possible value of λ:

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

Finally, we set the optimal λ_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.