Fifty Shades of Active and Index Alpha*

Markus Leippold[†] Roger Rueegg[‡]

March 16, 2018

Abstract

There is an ongoing debate about the benefits of active investing versus index investing. This paper will describe the different shades of alpha for active and index investing. It will differentiate across institutional and retail funds, before and after fees. It will also be benchmarked against one-factor, multi-factor and investable models, different markets, and different investment categories. We find evidence that the hypothesis of a zero-sum game after costs cannot be rejected for a vast majority of investment categories especially when we filter out the worst-performing mutual funds of the past year.

JEL classification: C12, C14, C22, G10, G14, G20, G21, G23

Keywords: active investing, index investing, mutual funds, robust alpha test

---

*We are indebted to Fabian Ackermann, Andreas Blöchlinger, Thorsten Hens, Andreas Kappler, Harald Lohre, and Michael Wolf for helpful comments. We thank the participants at the Frontiers of Factor Investing Conference 2018, and the seminar participants at the University of Zurich, Zurich Cantonalbank, and at the 2018 UBS Quant Conference in London.

[†]Department of Banking and Finance, University of Zurich, Switzerland, email: markus.leippold@bf.uzh.ch.

[‡]Department of Banking and Finance, University of Zurich and Zurich Cantonal Bank, Switzerland, email: roger.rueegg@zkb.ch.

1 Introduction

The emergence of index investing has led to the seemingly endless debate about the merits of active portfolio management. Many research papers, investors, and advisors place themselves in either the active or passive camp. The staunch defenders of active investing argue along the lines of Berk and Green (2004), who show that rational markets do not contradict the existence of skilled fund managers consistently beating the market. These scholars build their argument on a basic principle of economics: agents earn economic rents, if and only if they have a competitive advantage. Recently, Berk and van Binsbergen (2015) have provided empirical support for the claim that mutual fund managers do have skills.

In contrast, the proponents of passive investing argue along the lines of Fama and French (2010) in that the high fees of active management turn it into a negative-sum game after costs. Indeed, French (2008) and Fama and French (2010), among many others, provide ample evidence that actively managed US equity retail mutual funds underperform the US equity market portfolio after fees. Consequently, over the last couple of years, we have witnessed a massive inflow of funds into index investing. These observations naturally drive us to question the value of active management.

Our central message is that there is no black-or-white conclusion to the long-standing question of whether active management adds value. Indeed, there are many shades of gray when we compare active and index mutual funds, depending on the investor, the investment category, and the time period analyzed. Our paper contributes to the current literature in at least three aspects. First, we show the different shades of alpha for active and index investing. To achieve this goal, we differentiate across institutional and retail funds, we analyze performance before and after fees, we benchmark against one-factor and multi-factor models, and we distinguish between geographical regions and investment categories. For this extensive study, we include 61,269 funds across different asset classes.[1]

Second, our preliminary data analysis indicates that there is dependence over time in mutual fund returns. Therefore, for our performance analysis, we develop a robust test for the manager's alpha; that is, the excess return relative to an appropriate benchmark. Our test is robust in the sense that it accounts for potential serial dependence in mutual fund returns. Our method is able to to provide

[1] 56,136 of these funds are actively managed and 5,133 are index funds.

us with robust test statistics of the alpha, which we can then use as input for state of the art multiple hypothesis adjustments.

Third, we benchmark active managers not only against hypothetical factor portfolio returns but also against the recently emerged index funds. This is a more reasonable benchmark because it is an investable benchmark that includes all of the costs that arise when following a certain strategy, which is not the case when benchmarking against hypothetical factor portfolios. Furthermore, by using index funds, we avoid a potential selection bias in the benchmark model because the market capitalized benchmark of all of the index funds within a particular universe is known ex-ante. However, there is still an on-going debate about which multi-factor model best describes the investment opportunity set. By using index funds as a benchmark, we are able to select the average index investors' opinion instead of relying on ex-post views.

A robust statistical test for alpha helps us to achieve our goal, which is to better understand the impact of standard inference tests on the empirical rejection probabilities, while also controlling for multiple hypotheses. Consequently, we perform a simulation study. We find that the standard inference techniques are liberal in rejecting the null hypothesis, while we find accurate empirical rejection probabilities for our block resampling based alpha test. Other papers that conduct bootstrapped inference, such as Kosowski et al. (2006) and Fama and French (2010), sample one-period returns and, therefore, lose any information on the potential dependence over time. When we analyze the multiple hypothesis settings, we find evidence of cross dependence for connected time series of mutual fund portfolios. At the same time, there is less evidence for single mutual funds with a small overlap of returns. Therefore, for the serial and cross dependence, when we compare portfolios of mutual funds, we control the family-wise error rate by the method of Romano and Wolf (2005a,b, 2016). For a higher number of hypotheses in the single fund analysis, we regard the funds in isolation and control for multiple tries using the approach developed by Barras et al. (2010), which does not control for cross dependence.

When we apply the robust multiple hypothesis alpha test to portfolios of mutual funds and single mutual funds across different equity markets and then compare them with some commonly used multi-factor models, we find that active mutual funds deliver positive or close to zero alphas before fees and negative alphas after fees. Therefore, we provide international evidence for the results of Fama and

French (2010). However, when we apply the same tests to index funds for investment regions outside of the US, we find that index funds also show negative alphas after costs and that active mutual funds can even offer higher alphas for a majority of the analysis. While we agree with Fama and French (2010) in that US index funds deliver close to zero alphas, we conclude that this observation does not carry over to other investment regions. In particular, we find the highest proportion of skilled funds for institutional investors in Europe and Japan, which confirms the findings of Banegas et al. (2013).

In a next step, we compare the value-weighted returns of active against index mutual funds within the same investment category, as defined by Morningstar, sidestepping the need to choose appropriate multi-factor models as benchmarks. Such a direct comparison enables us to account for the transaction costs, restrictions, and management fees that arise in index strategies and means that we can reduce the selection bias of multi-factor models to a minimum. This exercise provides us with a fair and impartial comparison of active investing versus its investable index alternatives. Our analysis is based on 30 different investment categories across asset classes. We find that there are significant negative alphas after cost for the institutional 'US Equity Large Cap Blend' and retail 'Canada Fixed Income' categories. This finding corroborates the conclusion of Fama and French (2010).

However, for all of the other categories, our results support Berk and Green (2004) and Berk and van Binsbergen (2015) in that we cannot reject the hypothesis that it is a zero-sum game after costs. Hence, our findings resonate well with Pastor et al. (2015), who argue that in markets in which the mutual fund industry is big, such as the 'US Equity Large Cap Blend', active alphas tend to be negative and the equal-weighted alpha within investment categories exceed the value-weighted alpha. Furthermore, we see periods in which the average active managers underperform the index alternatives; such as, before the dot-com bubble burst, in the financial crisis, or in the very recent period from 2014 to 2016. But then we can also observe periods in which active managers, on average, outperform; such as, from 2000 to 2007 or from 2009 to 2014.

We find that after fees, there are a majority of unskilled mutual funds for the retail segment. In contrast, we see a more balanced proportion of skilled and unskilled funds for the institutional sectors and outside of the US. Thus, our results provide direct evidence for Garleanu and Pedersen (2015), who argue that more sophisticated investors outperform small investors because of the higher economies of scale in searching skilled active managers. Moreover, our results endorse the hypothesis of Gennaioli et al. (2015), who claim trust to be an essential component of high fees in asset management and who argue that active retail managers profit from pandering to trusting investors by buying hot assets, which explains the tendency for value-weighted retail mutual funds to have positive exposure to growth. Meanwhile, we find the opposite positioning in the institutional portfolios.

We further demonstrate that the average active retail investor can significantly improve her performance over the period ranging from 1993 to 2016 when the worst-performing active mutual funds of the past year are neglected. However, when the investor concentrates only on the top performing funds, the portfolio mainly loads on systematic risk and the overall performance cannot be significantly improved. In that sense, our finding provides further evidence for the results in Carhart (1997), who cannot explain the strong underperformance for the worst-return US retail mutual funds of the past year. As a consequence, when we exclude the worst-performing active mutual funds, we can increase the confidence for the zero-sum game hypothesis after costs.

The rest of this paper is structured as follows. Section 2 discusses the data wen we perform a preliminary analysis of the time dependence of mutual fund returns, which guides the design of our empirical tests. Section 3 presents a block resampling based robust alpha test. Using a set of simulation studies, we demonstrate the power of our robust alpha test. In Section 4, we compare the performance of index and active mutual equity funds when benchmarked against commonly used factor models. Section 5 provides a comparison of the value-weighted performance of active and index mutual funds across investment categories, and also in asset classes. We also show that active funds exhibit some persistence in returns. Section 6 concludes this paper. We delegate the data description to an appendix.[2]

2 Preliminary analysis

We first provide an overview of the data underlying our study. We then analyze the potential time dependency in mutual fund returns, which serves us a guidance for the formulation of an appropriate test statistics to evaluate our hypotheses.

[2]The Matlab code used in this paper is available from the authors on request.

2.1 Data

Our mutual fund sample is drawn from the Morningstar database and it ranges from December 1991 to December 2016.[3] In total, we include 61,269 funds across different asset classes. In comparison, Pastor et al. (2015) explore 3,126 actively managed US equity-only mutual funds while Berk and van Binsbergen (2015) use 5,974 actively managed funds. Hence, we add to the existing literature by providing evidence based on our new dataset. Furthermore, to the best of our knowledge, we are the first to apply a robust multiple hypothesis framework to active and index mutual funds in an international context. In Table 1 we provide the time-series averages of cross-sectional monthly attributes across asset classes. For the active funds, we analyze in total 14,969 institutional and 46,300 retail funds, while we have 56,136 active and 5,133 index funds. In general, there are fewer index funds, but they show higher average total net assets (TNA) and net returns, and also lower fees and about the same average years in the database. As expected, the institutional funds charge lower fees compared to their retail counterparts.

[Table 1 about here.]

In Figure 1, we plot the total net assets and the percentage of index fund within the dimensions equity and fixed income as well as retail and institutional funds. As of December 2016, the total net assets of equity retail funds ('EQR') amount to USD 9 trillion, those of fixed income retail ('FIR') funds to USD 3.7 trillion, and those of equity institutional funds ('EQI') and fixed income institutional funds ('FII') to USD 3.1 trillion and USD 2 trillion, respectively. Since institutional investors often invest their money by mandates, there are less institutional than retail funds. From the lower panel, we also observe that the percentage of index funds is steadily increasing. We find the highest concentration of index funds for equity funds, with 28% for retail and 32% for institutional funds. Looking at the fixed income funds, we find 18% of the retail and 13% of the institutional funds as being index funds.

[Figure 1 about here.]

[3] We provide a more detailed description of the data and data cleaning procedures in Appendix A.

2.2 Dependence over time

In their paper, Fama and French (2010) correctly point to a potential caveat in their resampling approach: because they perform a random sampling of months, they lose any effects of autocorrelation. However, they argue along the lines of Fama (1965) that this is a minor problem. In addition, Barras et al. (2010) do not account for potential dependency, claiming that they find such an effect only for a few mutual funds. For our data sample, we do not take these previous observations for granted. Therefore, we first take a closer look at our data. To test for autocorrelation in mutual fund returns, we apply two statistical tests. We start with the classical Ljung-Box (LJ) test statistic. However, the LJ test is based on autocorrelations and, hence, it is not a real test of independence.[4] Consequently, we use the test developed by Genest and Rémillard (2004) because it is based on ranks and, therefore, is distribution-free and does not depend on the underlying distribution of the observations.

Given that we will analyze the returns of single mutual funds and portfolios of mutual funds, and then compare them against multi-factor and investable benchmark models, we will apply the tests of time dependence to a variety of settings. We do not include all the funds in the dependency analysis, but we use a large representative sample with a special focus on the US market to particularly challenge the claim of Fama and French (2010). For the analysis of single mutual funds, we focus on all equity ('Equity') and fixed income ('Fixed Income') retail mutual funds and the 'US Equity Large Cap Blend' funds as defined by Morningstar ('US Equity'). For the portfolio returns, we value-weight all of the active retail mutual funds within the investment categories ('Inv. Categories'), where we find active and index mutual funds ranging back to 1993.[5] For the multi-factor comparison, we also build value-weighted portfolios of all mutual equity retail funds within the US, Global, Europe, Japan, and Asia ex-Japan investment regions ('Equity Regions').

To compare active and indexing investing, we construct two types of different benchmark models. First, we apply the the commonly used multi-factor models. In particular, we use the regional three-factor model ('FF3') with 'market', 'size', and 'value' factor as given in Fama and French (1992). We also use the five-factor model ('FF5') adding 'momentum' of Jegadeesh and Titman (1993) and the 'betting against beta' factor Frazzini and Pedersen (2010). Second, we focus on the investable

[4] One could easily construct a time series with zero autocorrelation while having very strong serial dependence.
[5] This categorization is called 'Global Category' within Morningstar.

one-factor benchmark model ('Investable'), which we build as the value-weighted return of the index funds within the investment category of the analyzed time series. For our dependency analysis, we compare the single 'Equity' and 'Fixed Income' funds, and also the 'Inv. Categories' portfolios with the investable benchmark model, while we compute the single 'US Equity' funds' and 'Equity Regions' portfolio residuals based on the 'FF3' and 'FF5' model. For the investable one-factor model, we require at least 12 monthly returns and for the multi-factor models at least 36 monthly returns.

[Figure 2 about here.]

The weakness of both of the above dependency tests is that we must fix the number of lags L. To set the number of lags, we make use of the automatic block-length selection for the dependent bootstrap of Politis and White (2004) and the correction of Patton et al. (2009). Their estimator regards the fastest possible rate of convergence based on the correlogram of the time series data. Figure 2 shows the optimal circular block length based on the settings that we describe above.[6] We observe that most mutual funds show an optimal block size of two or three for each of the datasets. For all settings, we find that 60% to 80% of all of the funds have an optimal block length ranging between one and three. Therefore, we set the lag L to three for the two tests of dependence over time and we then analyze the autocorrelation within the last three lags.

[Table 2 about here.]

In Table 2 we report the number and percentage of funds that show a p-value below 5% based on Fisher's p-value of Genest and Rémillard (2004) —which combines Cramér-von Mises test statistics— and the classical LJ test. Based on Fisher's p-value, we find that the percentage of rejected null hypothesis of no serial dependence over time increases from the left to the right, starting with 16% of the single equity mutual funds ('Equity') and reaching 40% for the 'Equity Regions' analysis. We observe the same pattern for the LJ test. In particular, for the 'Portfolio' approach there are on average one of three investment categories or regions where we can reject the null hypothesis.

When we go into more detail, we find that mutual funds with the longest available time series show a higher percentage of rejections. For example, for the 2% oldest 'Equity' and 'Fixed Income' single

[6] The optimal block length is ceiled to the next integer. We also performed the same analysis for institutional funds and we obtain similar results. Therefore, we do not include them here, but they can be obtained from the authors.

mutual funds, we find statistically significant serial dependence in 40% and 63% of the cases. For the 'Equity Regions' analysis, the investment regions where we find no evidence of time dependence is the US and Asia ex-Japan. In the 'Inv. Categories' analysis, we can reject the null based on the Fisher's p-value for the 'Canada Fixed Income', 'Canada Equity Large Cap', 'Emerging Markets Equity', 'US Equity Large Cap Blend', 'US Equity Large Cap Growth', and 'US Fixed Income' categories.[7]

Overall, our evidence shows that already in the first three lags, signs of serial dependence can be found in roughly every fifth single mutual fund, and every third mutual fund portfolio. Therefore, this evidence more than justifies the need to control for dependence over time when we analyze the alpha of single and portfolios of mutual funds against different benchmark models.

3 Methodology

Given the presence of substantial time dependency in the mutual fund data, as presented above, we propose to block bootstrap the alpha of a strategy to its benchmark returns. As outlined by Lahiri (2003), Haerdle et al. (2003), and Ledoit and Wolf (2008, 2011), the studentized block bootstrap improves the inference accuracy for dependent time series data compared to other methods. In addition, we obtain bootstrapped t-statistics and p-values, which we can use as an input to the latest multiple hypothesis frameworks as suggested by Romano and Wolf (2005a), Romano and Wolf (2005b), Romano and Shaikh (2006), Romano and Wolf (2007), and Barras et al. (2010).

3.1 Robust alpha testing

We closely follow Ledoit and Wolf (2008, 2011), who study the related problem of testing whether two Sharpe ratios or two variances are equal. Consider a fund with time-t return y_t and a set of K benchmark factor returns x_{tk}, $k = 1, ..., K$. A total of T returns are observed. We assume that these observations are generated by a stationary multivariate return distribution with mean vector μ and

[7] We do not report these more detailed results here, but they can be obtained by the authors.

covariance matrix Σ:

$$\mu = \begin{pmatrix} \mu_y \\ \mu_{x_1} \\ \vdots \\ \mu_{x_K} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_y^2 & \sigma_{yx_1} & \cdots & \sigma_{yx_K} \\ \sigma_{x_1 y} & \sigma_{x_1}^2 & \cdots & \sigma_{x_1 x_K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_K y} & \sigma_{x_K x_1}^2 & \cdots & \sigma_{x_K}^2 \end{pmatrix}, \tag{1}$$

with the observed means $\hat{\mu}$ and sample covariance matrix $\hat{\Sigma}$. By defining a vector $\mu_X = (0, E[x_1], \cdots, E[x_K])'$, we can express the fund's alpha as

$$\alpha = E[y] - \mu_X' \Sigma_{XX}^{-1} y_X, \tag{2}$$

with

$$\Sigma_{XX} = \begin{pmatrix} 1 & E[x_1] & E[x_2] & \cdots & E[x_K] \\ E[x_1] & E[x_1^2] & E[x_1 x_2] & \cdots & E[x_1 x_K] \\ \vdots & \vdots & & \ddots & \vdots \\ E[x_K] & E[x_K x_1] & E[x_K x_2] & \cdots & E[x_K^2] \end{pmatrix} \quad \text{and} \quad y_X = \begin{pmatrix} E[y] \\ E[x_1 y] \\ \vdots \\ E[x_K y] \end{pmatrix}, \tag{3}$$

Then, we test for the hypothesis

$$H_0: \quad \alpha = 0 \qquad \text{against} \qquad H_1: \quad \alpha \neq 0. \tag{4}$$

Furthermore, we define $\zeta_k = E[yx_k]$, $\gamma_k = E[x_k^2]$, $\xi_{kj} = E[x_k x_j]$, $j > k$, and the combined vector $\nu = (\mu_y, \cdots, \mu_{x_k}, \cdots, \zeta_k, \cdots, \gamma_k, \cdots, \xi_{kj}, \cdots)' \in \mathbb{R}^{1+3k+k(k-1)/2}$ with sample counterpart $\hat{\nu}$. Then, we can express the true alpha as a function f of ν, i.e.,

$$\alpha = E[y] - \mu_X' \Sigma_{XX}^{-1} y_X = f(\nu) \tag{5}$$

and the estimated alpha as function of $\hat{\nu}$, i.e., $\hat{\alpha} = f(\hat{\nu})$. As mentioned in Ledoit and Wolf (2008), under mild regularity conditions, it holds that

$$\sqrt{T}(\hat{\nu} - \nu) \xrightarrow{d} N(0, \Psi), \tag{6}$$

where Ψ is an unknown symmetric positive semi-definite matrix. By the delta method, we obtain

$$\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \nabla' f(\nu)\Psi\nabla f(\nu)) \tag{7}$$

with

$$\nabla' f(\nu) = \left(\frac{\partial f(v)}{\partial \mu_y}, \cdots, \frac{\partial f(v)}{\partial \mu_{x_k}}, \cdots, \frac{\partial f(v)}{\partial \zeta_k}, \cdots, \frac{\partial f(v)}{\partial \gamma_k}, \cdots, \frac{\partial f(v)}{\partial \xi_{kj}}, \cdots \right)'. \tag{8}$$

Given a consistent estimator $\hat{\Psi}$ of Ψ, we can compute a standard error for $\hat{\alpha}$ by

$$s(\hat{\alpha}) = \sqrt{\frac{\nabla' f(\nu)\hat{\Psi}\nabla f(\nu)}{T}}. \tag{9}$$

To test the null hypothesis in Equation (4), we focus on the bootstrap inference for time series data outlined in Ledoit and Wolf (2008). In particular, we denote the optimal block length with b and define $l = \text{floor}(T/b)$. As shown in Kuensch and Goetze (1996), the bootstrapped estimator of $\hat{\Psi}^*$ is given by

$$\hat{\Psi}^* = \frac{1}{l}\sum_{j=1}^{l} \eta_j \eta_j', \tag{10}$$

where

$$\begin{aligned} z_t^* &= \left(y_t^* - \hat{\mu_y}^*, ..., x_{tk}^* - \hat{\mu_x}^*, ..., y_t x_{tk} - \hat{\zeta}_k^*, ..., {x_{tk}^*}^2 - \hat{\gamma}_k^*, ..., x_{tk}^* x_{tj}^* - \hat{\xi}_{kj}, ... \right), \\ \eta_j &= \frac{1}{\sqrt{b}} \sum_{t=1}^{b} z_{(j-1)b+t}. \end{aligned} \tag{11}$$

The bootstrapped standard error $s\left(\hat{\alpha}_m^*\right)$ of the m-th bootstrap $m = 1, ..., M$ is then computed as in Equation (9). Next, the studentized statistics are built as

$$\tilde{d}_m^* = \frac{|\hat{\alpha}_m^* - \hat{\alpha}|}{s\left(\hat{\alpha}_m^*\right)}, \tag{12}$$

and the p-value is calculated as

$$PV = \frac{\{\tilde{d}_m^* \geq \hat{d}\} + 1}{M + 1}, \tag{13}$$

where $\hat{d}$ is the original studentized test statistic that was computed from the observed returns. We use Newey-West standard errors to calculate the original standard errors. Regarding the optimal block, we either suggest to use the method of Politis and White (2004) and Patton et al. (2009) for the univariate case, or the method of Ledoit and Wolf (2008) for the bivariate case. For our empirical analysis, we would like to compare up to 30 different investment categories and, so far, there is no available method to make this comparison. Consequently, we will further discuss the optimal block size to use in our simulations in Section 3.3.

3.2 Multiple Hypothesis Testing

To account for potential time dependencies in the funds' alpha in a multiple hypothesis testing framework, we will use the bootstrapped studentized statistics and p-values in Equations (12) and (13).[8] When controlling for multiple tries, we will either use the family-wise error rate (FWER) or the false discovery rate (FDR). The FWER dates back to the work of Bonferroni (1936) and it is defined as the probability of at least one false discovery. Romano and Wolf (2005a,b) introduce a stepwise multiple testing procedure that not only gives higher statistical power than the tests of Bonferroni (1936) and Holm (1979) but also allows for cross-sectional dependence. Romano and Wolf (2016) provide an efficient method to calculate the adjusted p-values without the need to fix a prespecified significance level.

The FDR is defined as the expectation of the proportion of falsely rejected null hypotheses. For a larger number of hypotheses and under dependence, Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001) show that it is favorable to control for the FDR. Meanwhile, Barras et al. (2010) introduce a testing framework that controls the FDR for the analysis of mutual funds. In addition, Bajgrowicz and Scaillet (2012) argue that investors do not rely on a single active manager but instead diversify between different managers. Therefore, in their view, it is favorable to control the amount of falsely rejected hypotheses instead of investing only in the best strategies, as is the case for the more conservative FWER.

In this study, we would like to test whether single active or index funds significantly outperform

[8] As we have shown in our simulation exercise, even if there is no serial dependency, our alpha test statistic is accurate.

the theoretical multi-factor models. For this single mutual funds' multiple hypothesis problem, we have a large number of hypotheses and thus prefer to control for the FDR. We also observe only a few overlapping time series, which makes the analysis of the cross-dependence difficult. While Fama and French (2010) overcome this problem by jointly sampling the funds' and factors' returns at only one point in time, we cannot jointly sample blocks of time series where we have missing values. The benefit of Fama and French (2010) is that they take into account the cross-correlation in the test statistics. However, the cost is that they lose any dependence over time and the number of months of a fund in a simulation depends on its lifespan.

Because Barras et al. (2010) only find an average pairwise correlation of 0.08 in their sample, we also analyze the cross-sectional dependence of our data. For this analysis, we reuse the single fund settings 'US Equity' with the 'FF3' and 'FF5' residuals, and the 'Equity' and also 'Fixed Income' with the investable one-factor model from Section 2. To detect cross-sectional dependence, we apply the test of Pesaran (2004). Compared to the well-known Lagrange multiplier test of Breusch and Pagan,[9] this test is correctly centered for a large sample and comparably short time series. This can be observed in our case with a large cross-section but a comparably small time series. To compute the test statistic, we concentrate on funds with more than 14 years of data.[10] For the three-factor and five-factor residuals, and for the 'US Equity' data, we find a p-value of 0.23 and 0.22, while for the one-factor residuals we find a p-value of 0.47 and 0.5, respectively, for the 'Equity' and 'Fixed Income' data. Consequently, we cannot reject the hypothesis of no cross-sectional dependence and we prefer to apply the robust alpha test in combination with the approach of Barras et al. (2010) to estimate the percentage of skilled and unskilled funds. This allows us to regard each fund in isolation and accounts for both the evidence of time dependence from Section 2 and multiple tries.

For the analysis of portfolios of mutual funds, we have only a few hypotheses. Therefore, we prefer to control the FWER because the diversification argument of Bajgrowicz and Scaillet (2012) does not hold anymore. Because of the formation of portfolios, we also have no missing values. In addition, we can jointly sample blocks of funds' and benchmark models' returns and sample the test statistics with the robust alpha test from the previous section. Based on the sampled test statistics that

[9] See, Breusch and Pagan (1980).

[10] This means that we focus on the 'US Equity' data on 42%, the 'Equity' data on 34%, and the 'Fixed Income' data on 27% of the available funds.

account for time dependence, we can next apply the state of the art multiple hypothesis framework of Romano and Wolf (2016) that also controls for cross-dependence of the hypotheses. This caution is important because when we apply the test of Pesaran (2004) to the portfolio settings from Section 2 we can reject the hypothesis of no cross-correlation for the 'Inv. Categories' residuals at the 5% significance level. Meanwhile, in this test there is no evidence for the 'Equity Regions' analysis. However, because we have a small cross-section but a comparably large time series, we can also apply the well-known Lagrange multiplier test of Breusch and Pagan, which rejects the null hypothesis of no cross-dependence for all portfolio settings 'Inv. Categories' and 'Equity Regions'.

3.3 Simulation Study

We now perform a simulation study of the approach that was presented in the previous section. For this purpose, we first simulate a single hypothesis setting. For realistic time series, we select the first 10 US mutual funds of the Morningstar database within the category 'US Equity Large Cap Blend' that offer the entire return history from 1992 to 2016 (T=300). For their benchmark model, we focus on the one-factor 'CAPM', three-factor 'FF3', and five-factor 'FF5' model. For the data generating process (DGP), we sample from the realized returns with a circular block bootstrap and block sizes of 1, 3, and 6. We select this grid of block sizes based on our analysis in Section 2, where we observe that most of the optimal circular block sizes range between 1 and 6. The grid corresponds to the time periods of one, three, and six months. The block sizes 3 and 6 are the ones that take the evidence of serial dependence from Section 2 into account. The block size 1 generates independent data, and we employ this block size for comparison reasons only. For each fund, we simulate 1,000 paths and set the alpha under the null hypothesis to the true observed alpha of the data. The bootstrapped p-values ('Boot') are then calculated as illustrated in Section 3.1 by employing M = 1,000 and the optimal block size by the method of Politis and White (2004) and the correction of Patton et al. (2009). We compare the robust p-values against the standard inference methods; that is, based on the normal distribution (Standard), Newey-West (NW), and HC3 standard errors.

[Table 3 about here.]

In Table 3 we report the empirical rejection probabilities of the falsely rejected null hypothesis compared to the nominal levels $\alpha = 10\%$, $\alpha = 5\%$, and $\alpha = 1\%$. Because the null hypothesis is true for all the simulations, the true rejection probabilities should be equal to the nominal levels of the test. If a test shows a higher percentage of rejections, then we regard this test as too liberal. While we observe that the standard inference tests based on the normal distribution, the Newey-West, and HC3 standard errors are liberal in rejecting the null hypothesis, the bootstrapped solution ('Boot') presented in the previous section is more conservative and closer to the nominal levels. We highlight in bold the empirical rejection probabilities that are closest to the desired level. We only observe the HC3 standard errors to be in some cases closer to the desired level compared to the block bootstrapped method but only in the case where we apply the standard but less realistic bootstrap with a block size of one. However, as we demonstrate in Section 2, the optimal block size and thus a realistic assumption for the DGP is, in general, around three or six, where the bootstrapped test presented in this work is more accurate as also observed in Ledoit and Wolf (2008) for the Sharpe Ratio and Ledoit and Wolf (2011) for the variance.

Since there is still the open question of the optimal block size in the multiple hypothesis setting when controlling the FWER as illustrated in Romano and Wolf (2005a,b, 2016), we conduct a second simulations study. Unlike in the single mutual funds' analysis, where we regard each fund in isolation and then apply the multiple hypothesis framework of Barras et al. (2010), in this case, we must consider the cross-dependence structure, and jointly sample the funds and benchmark returns. For this purpose, we focus on the 17 portfolios within the 'Inv. Categories' setting from Section 2 with the investable one-factor benchmark model that is based on the value-weighted return of index funds. Also instead of calculating the Type I Errors as in the single hypothesis setting, we compute the empirical rejection probabilities based on the FWER; as illustrated in Romano and Wolf (2005a,b, 2016). To find the optimal block size that is closest to the nominal levels of the test we focus on the following grid of block sizes: 1, 3, 6, 9, and 12. Regarding the DGP we keep the grid from our first simulation study.

[Table 4 about here.]

We show the empirical rejection probabilities based on the FWER in Table 4. Likewise, for the FWER, we find the bootstrapped robust alpha test to achieve the desired levels at optimal block sizes three, or six. Given that we observe for a block size of three accurate rejection probabilities, in the multiple comparisons of portfolios, we will report the results based on the optimal block size of three in the remainder of the paper. Finally, the more conservative block sizes six and nine are applied for robustness checks.

4 Multi-factor benchmark models: International evidence

Although previous studies have focused on active retail US mutual funds, the performance analysis of index mutual funds or other investment regions is often omitted. Ferreira et al. (2013) published two international studies for the mutual funds' performance, while Banegas et al. (2013) concentrates on the characteristics of European Mutual Funds. Ferreira et al. (2013) finds that mutual funds around the globe, in general, underperform. In contrast, Banegas et al. (2013) find that European funds outperform the market.

Because the choice of the replicating index is an active decision, it is of interest whether index funds exhibit style exposures. Therefore, we first analyze the portfolio of value-weighted alphas of active and index equity funds across different investment regions. Later, we compare the distribution of the single mutual funds' alphas' robust p-values against the regional multi-factor benchmark models. For the factors, we apply the three-factor model 'market', 'size' and 'value', as in Fama and French (1992), and also the five-factor model, including the 'momentum' and 'betting against beta' factor of Jegadeesh and Titman (1993) and Frazzini and Pedersen (2010).[11]

4.1 Portfolios of Mutual Funds

We first form portfolios according to the investment regions 'US Equity', 'Global Equity', 'Europe Equity', 'Japan Equity,' and 'Asia ex-Japan Equity.' We then benchmark them against the regional multi-factor models. Many investors are unaware that index investing is active investing in the sense that investors have to select a replicating index actively. This means that they are not acting as a passive investor in the view of the Capital Asset Pricing Model (CAPM), as developed by Sharpe

[11]The regional returns for the 'market', 'size', 'value', and 'momentum' factors are retrieved from the homepage of Fama and French (1992), while the 'betting against beta' factor is provided for each region on the homepage of Frazzini and Pedersen (2010).

(1964) and Lintner (1965). As illustrated by Roll (1977), the equilibrium itself is not observable and, therefore, the index is only an approximation of the market portfolio. Following Berk and Green (2004) and Berk and van Binsbergen (2015), we value-weight the alpha of each fund within its region.

[Table 5 about here.]

In Table 5, we report the results of our regression analysis for retail mutual funds based on the value-weighted portfolio returns and the multi-factor models. The returns of the mutual funds are based on net returns, while the benchmark factors are gross of fees. Therefore, and not surprisingly, we find negative alphas for both active and index funds. We highlight two important observations. First, we show a significant small (large) cap bias for active (index) mutual funds. Second, we find often significant positive exposures to the value factor (HML) for the aggregated index funds.

[Table 6 about here.]

We observe similar patterns in the same analysis for institutional funds. In Table 6, the alphas of the active funds are in more than half of the benchmark models and regions higher compared to index funds. In addition, active institutional funds show a positive sensitivity to the value factor, in contrast to the active retail funds in most of the regions, where we often find negative exposures to the value factor. Asia ex-Japan is one exception, where we see a significant positive exposure to value for both active retail and institutional mutual funds. The higher exposure of retail funds to growth stocks adds evidence to the theory of Gennaioli et al. (2015) that active managers profit from pandering to trusting investors when they invest in hot assets such as growth stocks.

[Table 7 about here.]

We now apply the newly created robust alpha test to the different investment regions and adjust for multiple hypotheses. In Table 7, we show the results of the one-, three-, and five-factor models including the same factors as outlined in the regression above. We report the multiple hypothesis adjusted p-value of Romano and Wolf (2016) as outlined in Section 3.2.[12] We report the p-values

[12] Because the number of hypotheses is small we control the well-known FWER.

based on an optimal block size of three as discussed in Section 3.3. We highlight multiple hypothesis adjusted p-values below the 10% significance level. When we apply the block sizes six and nine, the p-values increase slightly, but the significant p-values stay the same except for the significant negative p-values for the retail European funds. The analysis period starts in February 1998 and it ends in December 2016. This is the period where we have return data across all investment regions for index and active mutual retail funds. For institutional funds in Japan and Asia ex-Japan, the data points do not reach as far back and, therefore, we exclude them from our analysis.

When we focus on the active mutual funds, we find, on the one hand, and in line with the results of Fama and French (2010), close to zero alphas before fees and negative alphas after fees for US Equity funds, but also other investment regions. For the US and Global institutional funds in the three-, and five-factor model, as well as for the one-factor model and retail funds in Europe we find that they are significantly different, even when we adjust for multiple tries. On the other hand, and as suggested by Berk and Green (2004), we conclude for all the other benchmark model and regions that the value-weighted alpha is not different from zero. Thus, also for the one-factor model and active retail funds within the US, we cannot reject the null hypothesis of a zero-sum game after costs at the significance level as high as 10%. We even find investment regions where the net alpha is positive. Surprisingly, for European institutional funds, we see a significant positive alpha for institutional mutual funds before fees. Thus, we can confirm the results of Ferreira et al. (2013), which also find skill in Europe.

When we now focus on the alpha of the index funds, we also find mostly negative alphas before and after costs. Fama and French (2010) mention in their appendix that they estimate the alpha of a value-weighted portfolio of index funds for the one-, three- and four-factors model to be 0.08%, -0.16%, and 0.01% per year. Although our results confirm their findings for the US, for the other investment regions 'Global', 'Japan' and 'Asia ex-Japan' we find for 13 out of the 18 comparisons higher net alphas of active mutual funds compared to index mutual funds. For the 'Global Equity' funds, we even observe significant negative alphas before and after fees. We explain this by the fact that for global indices the choice of the index is the most heterogenic and, therefore, most crucial among the investment regions. There are index funds that overweight specific investment regions due to a possible home bias. Or, there are thematic or smart beta funds where the weights of the replicating can arise due to bottom-up security characteristics. The less observable or, the more heterogenic the view about the replicating index, the more index investing turns to active investing and distances itself from the passive investor; as described in the CAPM model.

4.2 Single mutual funds

Next, we apply the robust alpha test to each of the individual funds' net returns with the same multi-factor benchmark models. In Figure 3, we show the histograms of the robust p-values for retail and institutional mutual funds. We only report the results for the five-factor model because the other models show similar patterns. For each individual fund, we apply the optimal block size with the method of Politis and White (2004) and Patton et al. (2009).[13]

[Figure 3 about here.]

The patterns support our previous findings. First, we find a higher number of mutual funds with a negative and significant alpha. For institutional funds, the number of significant negative and positive alphas is more balanced. The positive alphas of European institutional funds of Table 7 are the result of a higher than expected number of significant outperforming funds in the right tail of the distribution. Surprisingly, we also observe the same pattern in the tails of the index mutual funds. There are more index funds with significant negative p-values. In addition, the ranking and the percentage of significant underperforming funds is very similar across the investment regions for the index and active funds. When we compare the equally weighted histograms on the left with the total assets over time weighted analysis on the right, we find that for active funds there is a higher percentage of outperforming funds. While with equal weighting there is always a higher percentage of funds with significant negative p-values at the 5% level compared to positive p-values, although this is no longer the case for the value-weighted analysis. However, for the institutional funds in Japan and Europe, there is total asset-weighted a higher percentage of significant positive p-values, but this is not the case for the US. This confirms the observation of Ferreira et al. (2013) who find—in contrast to the US—that mutual funds from overseas are less affected by diminishing returns to scale.

[13] For robustness, we also applied a block size of six, which shows the same results.

We now correct for multiple tries by the method of Barras et al. (2010) as described in Section 3.2 and compute the proportion of zero alpha, unskilled, and skilled mutual funds. We report the results in Table 8 for the robust p-values based on the five-factors benchmark model.

[Table 8 about here.]

We find that the proportion of zero alpha active funds is 62.3% for retail and 77.4% for institutional funds. As expected the portion of zero alpha funds for index investors is higher at 68.4% for retail and 79.5% for institutional funds. We also find that the percentage of skilled funds is the highest for active institutional funds with 3.5%, while for all the other categories there are between 1.4% and 1.9% skilled single mutual funds. The proportion of unskilled funds is the highest for active retail funds with 35.9%, while active institutional funds show 19.1% of unskilled managers. We also find that 29.7% of the retail and 19.1% of the institutional index funds are unskilled. This is further evidence that even index investors run the risk of selecting an unfavorable index mutual fund. For the active and index mutual funds, we only find skilled managers in Europe and Japan, while for the other markets we detect zero skilled managers when we correct for multiple tires. For institutional investors in Europe and Japan, the proportion of skilled to unskilled managers is the most balanced with 8.2% to 13.3% in Europe and 9.4% to 2.4% in Japan. When we conduct the same analysis for the one-factor model, we find the same patterns except that the skilled active managers in Japan disappear. However, we must treat these results with care because single fund multiple hypothesis analysis does not take the fund size into account.

5 Comparison of active vs index investing

Fama and French (2010) argue that the comparison of active versus index investing leads to the same results as comparing active performance with the multi-factor models because the value-weighted portfolio of index funds exhibits close to zero alphas. Although we agree with their arguments for the US, we demonstrated in the previous section that index investing exhibits negative alphas in other equity regions. Therefore, whether the two true alternatives active or index investing differ across asset classes and across different markets is still open to question. In our view, there are at least three arguments that the value-weighted return of active and index funds within the same Morningstar investment category builds a fair comparison. First, multi-factor models include long-short portfolios with, like for momentum, very high turnovers. In practice, transaction costs occur and more importantly shorting of securities is costly and only possible under certain short-selling restrictions. In addition, multi-factor models are not equilibrium models. This means that we don't know ex-ante the optimal allocation in the factors and even more crucial not every investor can invest in the same portfolio of factors. Then, for every factor investor, there must be another investor in the market that takes the opposite side of the trade. The value-weighted return includes all transaction cost and restrictions, and it also takes into account that factors with a low market capitalization obtain a lower weight.

Second, multi-factor benchmark models make active decisions. There is no general agreement on the numbers and kind of anomalies one should include in a multi-factor benchmark model. This leads to a severe selection bias, which will also be suffered by index investors. In contrast, the value-weighted portfolio of index funds is free of this selection bias and it only includes the factor returns when there exist index funds that track this kind of strategy at a specific point in time.

Third, in practice, when investors or active managers focus on a specific investment category, they do not compare themselves with the multi-factor models in general but they do instead compare themselves with other funds within the same category. Morningstar itself corrects for biases when they build categories such as small cap, growth or value groups. Thus, when we apply the value-weighted return within an investment category, we compare the average active investor with the average index investor that decided to invest in the same market.

It is of note that active mutual funds commonly report their returns compared to a self-selected benchmark. When we compare active mutual funds with the value-weighted portfolio of index funds, we disregard their benchmark. Given that we intend to analyze the performance of active versus index investing, this approach is more appropriate than a comparison to a hypothetical multi-factor benchmark. However, it may lead to situations in which active managers could beat (underperform) their benchmark but, due to their benchmark choice, possibly still underperform (lose) versus the value-weighted portfolio of index funds.

5.1 Portfolios of mutual funds

We first form value-weighted portfolios of active and index funds within the Morningstar categories and we also distinguish between institutional and retail funds. We regard the periods from 1993 to 2016 and 2000 to 2016, which allows us to include more investment categories for the more recent time periods. Given that index mutual funds only emerged recently, we observe at least one index fund for the longer collected data history for four institutional and 17 retail categories. For the more recent period starting in 2000, we find 30 and 12 investment categories for the retail respectively institutional segment.

[Figure 4 about here.]

Figure 4 shows the *p*-value adjusted for multiple hypothesis on the y-axis and the net and gross alpha on the x-axis. We plot for each of the available categories the results of our multiple hypothesis tests based on the method of Romano and Wolf (2016). After fees, we find the 'US Equity Large Cap Blend' category for institutional funds and the 'Canada Fixed Income' category for retail funds to significantly underperform the alternative of the value-weighted index funds for both periods. For the negative alpha of the 'Euro Fixed Income' retail category and the period from 1992 to 2016 we also find a significant *p*-value. Before fees, we find 'US Fixed Income' from 1992 to 2016, and 'Global Equity Large Cap', 'Emerging Markets Equity', and 'Europe Equity Large Cap' from 2000 to 2016 for institutional, and also 'Global Equity Large Cap' from 1992 to 2016 and 'Global Equity' from 2000 to 2016 for retail funds to significantly outperform the value-weighted index funds. 65% (44%) of the investment categories before (after) fees show a positive alpha over time. Thus, we find evidence for both theories. In a few categories we see a significant underperformance indeed, but for the majority of the categories, we cannot reject the hypothesis of a zero-sum game.

When we apply a block size of six or nine, the *p*-values increase slightly. The 'Euro Fixed Income', 'Emerging Markets Equity' and 'Europe Equity Large Cap' categories, that exhibit a *p*-value just below 10% for the block size of three, start to show insignificant *p*-values.

[Figure 5 about here.]

We next analyze the aggregated performance of the active minus the index value-weighted portfolio. We show in Figure 5 that we find that the aggregated performance of active mutual funds beats the aggregated performance of index funds for the institutional and retail funds the cumulated logarithmic performance when we equal-weight ('EW'), and when we value-weight the investment categories ('VW'). Before costs ('Gross'). After costs ('Net'), there is only the equal-weighted portfolio that shows a positive cumulated return over time. There are three major periods where active managers underperformed their index counterpart: before the burst of the dot-com bubble, in the financial crisis, and in the recent past. When we apply the robust alpha test, we find that except for the value-weighted institutional portfolio before fees, we cannot reject the hypothesis that the alpha is different from zero. The fact that institutional mutual funds' alpha is on aggregate higher compared to the retail segment adds evidence to the theory of Gennaioli et al. (2015) and Garleanu and Pedersen (2015) in the sense that institutional managers are more sophisticated and larger. Thus, trust plays a less important role, and also the search costs for skilled managers are lower relative to capital what reduces their fees and increases the alpha compared to index investing.

5.2 Single mutual funds

We now focus on the single p-values of our robust alpha test. Figure 6 shows the distribution of the p-values for active (index) mutual funds compared to the value-weighted index (active) benchmark in the top (bottom) chart. We find a higher percentage of funds with a negative net alpha for both active and index mutual funds. The percentage of significant negative mutual funds is also larger compared to the percentage of significant positive managers.

[Figure 6 about here.]

When we compute the percentage of skilled and unskilled funds using the method developed by Barras et al. (2010), we find for both index and active mutual funds nearly the same portion of skilled, unskilled and zero alpha funds. In Table 9 we show the percentage for each asset class (equity and fixed income) and segment (retail and institutional). We find that on average 78.7% for active and 82.9% for index mutual funds are zero alpha funds. The average percentage of skilled funds is the same 4%, while there are more unskilled funds in the active segment with 17.2% versus 12.8% for index funds.

[Table 9 about here.]

5.3 Persistence analysis

Carhart (1997) finds that US mutual funds with a substantial underperformance over the past year persist to underperform over the next year. In contrast to the outperformance of the best mutual funds, he cannot explain the persistence in the worst mutual funds. Thus, it is of interest whether an active investor is better-off if he avoids the losers of the past year. To simulate the return of an average active investor that trades according to this simple rule, we proceed as follows. Every year in December, we first sort the funds within each investment category based on the performance over the past year. Next, we invest in the asset-weighted portfolio of the x% best funds. We normalized the weights each month. Thus, if a fund shows no data for one month, it disappears from the portfolio. For x we chose steps of 10% starting with all mutual funds to the best 10% mutual funds. To aggregate the performance numbers of the different investment categories, we value-weight the net returns by the total active assets as in Figure 5 for the value-weighted portfolios. We disregard data points where we have less than 10 active mutual funds. We compare the persistence portfolios first with the performance of the value-weighted return of index funds and the value-weighted return of all active funds. We report the various performance numbers after fees of this analysis from 1993 to 2016 in Table 10.

[Table 10 about here.]

When we first compare the performance against the benchmark of the value-weighted index mutual funds in the top of the table, we find that the performance increases the more we filter out the worst performing funds. The performance difference from 1993 to 2016 climbs from 0.08% bps to 1.2% and -0.57% to 0.23% for institutional respectively retail mutual funds. For the alpha, we find a similar pattern. However, since we also observe higher betas for the portfolios with only the best mutual funds, a significant part of the performance is explained by the higher risk. When we regard the robust p-value of the test presented in this work we find that the p-value tends to increase from 0.44

for institutional and 0.18 for retail mutual funds to close to one. Thus, if an active investor selects only from the best mutual funds over the past year, the threshold to reject the zero-sum game after costs hypothesis increases.

Also, we find in the bottom of the table that the more we concentrate on the best mutual funds, the higher is the performance difference and alpha compared to a portfolio with all active mutual funds. But when we regard the p-values of the alpha, we only find significant values for the retail mutual funds up to a threshold of 30%. Thus, we see similar to Carhart (1997) that the one-year persistence is just significant for the underperforming retail mutual funds. The more we concentrate on the top performing mutual funds of the past year, the more of the performance can be explained by the higher systematic risk of these funds.

6 Conclusion

In this paper, we provide a performance analysis of traditional active mutual funds and the recently emerging index funds. Because we find that serial dependence might obscure the statistical analysis when using previous methods, we first create a robust alpha test. This test takes into account time dependence of returns and offers the basis for state-of-the-art multiple hypothesis tests. We find that the block resampling method provides high statistical power, while standard inference tests are liberal in rejecting the null hypothesis.

When we apply the robust alpha test to a rich data set of Morningstar covering 61,269 mutual funds across different regions and asset classes from 1992 to 2016 and compare the returns against common multi-factor models, we find no black-or-white conclusion. Instead, many different shades of gray emerge, depending on the investment category and benchmark model. We see, as in Fama and French (2010), that active equity managers show zero alphas before fees and negative alphas after fees. However, only for the US and Global institutional funds do we find a significant underperformance. For all the other markets and also for US retail funds, we cannot reject the hypothesis of a zeros sum game after costs. Thus, we find evidence for the theory of Berk and Green (2004). Particularly for Europe, we see that active managers do have skill. Hence, our results add to the work of Banegas et al. (2013).

Outside of the US, we also find that index funds underperform multi-factor benchmark models. The index alphas are even lower for a majority of the analyzed markets compared to active mutual funds. Thus, the argument of Fama and French (2010) that index investors show close to zero alphas is only valid in the US. For a fair comparison of active versus index investing, we compare the value-weighted return of active versus index mutual funds across the Morningstar investment categories. This also offers the advantage that we compare the performance of active managers against an investable benchmark model that is prone to transaction costs, management fees, and also reduces the benchmark selection bias to a minimum.

When we conduct the fair comparison across equity and fixed income mutual funds, we find significant negative alphas after fees for the 'US Equity Large Cap Blend' for institutional funds and 'Canada Fixed Income' for retail funds. But for the vast majority of the categories, we cannot reject the hypothesis of a zero-sum game after costs. Indeed, we even find categories such as 'US Fixed Income' and 'Global Equity Large Cap' for institutional investors with significant p-values before fees. Also, when an investor excludes the worst-performing retail mutual funds of the past year, she can significantly improve her performance over time. Thus, following this trading rule further strengthens our hypothesis of a zero-sum game after fees for active and index investing.

Overall, and when we analyze single mutual funds in isolation and the value-weighted portfolios of mutual funds, we find that institutional investors perform better compared to retail funds and that before fees active managers outperform, while after fees there is a majority of mutual funds or investment categories with negative performance. These findings confirm the theories of Gennaioli et al. (2015) and Garleanu and Pedersen (2015), which assign trust and the economies of scale in searching active managers a critical role in delegated asset management.

References

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

Banegas, Ayelen, Ben Gillen, Allan Timmermann, and Russ Wermers, 2013, The cross section of conditional mutual fund performance in European stock markets, *Journal of Financial Economics* 108, 699–726.

Barras, Laurent, Olivier Scalliet, and Russ Wermers, 2010, False discoveries in mutual fund performance: Measuring luck in estimated alphas, *Journal of Finance* 65, 179–216.

Benjamini, Yoav, and Yosef Hochberg, 1995, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society. Series B (Methodological)* 289–300.

Benjamini, Yoav, and Daniel Yekutieli, 2001, The control of the false discovery rate in multiple testing under dependency, *Annals of Statistics* 1165–1188.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Berk, Jonathan B., and Jules H. van Binsbergen, 2015, Measuring skill in the mutual fund industry, *Journal of Financial Economics* 118, 1 – 20.

Bonferroni, Carlo E., 1936, *Teoria statistica delle classi e calcolo delle probabilita* (Libreria internazionale Seeber).

Breusch, T. S., and A. R. Pagan, 1980, The lagrange multiplier test and its applications to model specification in econometrics, *Review of Economic Studies* 47, 239–253.

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

Elton, Edwin J., 2001, A first look at the accuracy of the crsp mutual fund database and a comparison of the crsp and morningstar mutual fund databases, *Journal of Finance* 56, 2415–2430.

Fama, Eugene F, 1965, The behavior of stock-market prices, *Journal of Business* 38, 34–105.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427.

Fama, Eugene F., and Kenneth R. French, 2010, Luck versus skill in the cross-section of mutual fund returns, *Journal of Finance* 65, 1915–1947.

Ferreira, Miguel A., Aneel Keswani, António F. Miguel, and Sofia B. Ramos, 2013, The determinants of mutual fund performance: A cross-country study, *Review of Finance* 17, 483–525.

Frazzini, Andrea, and Lasse H. Pedersen, 2010, Betting against beta, Technical report.

French, Kenneth R, 2008, Presidential address: The cost of active investing, *Journal of Finance* 63, 1537–1573.

Garleanu, Nicolae B, and Lasse H. Pedersen, 2015, Efficiently inefficient markets for assets and asset management, Technical report, National Bureau of Economic Research.

Genest, Christian, and Bruno Rémillard, 2004, Test of independence and randomness based on the empirical copula process, *Test* 13, 335–369.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, 2015, Money doctors, *Journal of Finance* 70, 91–114.

Haerdle, Wolfgang, Joel Horowitz, and Jens-Peter Kreiss, 2003, Bootstrap methods for time series, *International Statistical Review* 71, 435–459.

Holm, Sture, 1979, A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics* 65–70.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White, 2006, Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis, *Journal of Finance* 61, 2551–2595.

Kuensch, Hans Rudolf, and Friedrich Goetze, 1996, Second-order correctness of the blockwise bootstrap for stationary observations, *The Annals of Statistics* 24, 1914–1933.

Lahiri, Soumendra N., 2003, *Resampling methods for dependent data* (Springer New York).

Ledoit, Oliver, and Michael Wolf, 2008, Robust performance hypothesis testing with the Sharpe ratio, *Journal of Empirical Finance* 15, 850–859.

Ledoit, Olivier, and Michael Wolf, 2011, Robust performances hypothesis testing with the variance, *Wilmott* 2011, 86–89.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 13–37.

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2015, Scale and skill in active management, *Journal of Financial Economics* 116, 23 – 45.

Patton, Andrew, Dimitris N. Politis, and Halbert White, 2009, Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White, *Econometric Reviews* 28, 372–375.

Pesaran, Hashem M., 2004, General Diagnostic Tests for Cross Section Dependence in Panels, CESifo Working Paper Series 1229, CESifo Group Munich.

Politis, Dimitris N., and Halbert White, 2004, Automatic block-length aelection for the dependent bootstrap, *Econometric Reviews* 23, 53–70.

Ratcliff, John W., and David E. Metzener, 1988, Pattern-matching-the gestalt approach, *Dr Dobbs Journal* 13, 46.

Roll, Richard, 1977, A critique of the asset pricing theory's tests part i: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129 – 176.

Romano, Joseph P., and Azeem M. Shaikh, 2006, Stepup procedures for control of generalizations of the familywise error rate, *The Annals of Statistics* 34, 1850–1873.

Romano, Joseph P., and Michael Wolf, 2005a, Exact and approximate stepdown methods for multiple hypothesis testing, *Journal of the American Statistical Association* 100, 94–108.

Romano, Joseph P., and Michael Wolf, 2005b, Stepwise multiple testing as formalized data snooping, *Econometrica* 73, 1237–1282.

Romano, Joseph P., and Michael Wolf, 2007, Control of generalized error rates in multiple testing, *The Annals of Statistics* 35, 1378–1408.

Romano, Joseph P., and Michael Wolf, 2016, Efficient computation of adjusted p-values for resampling-based stepdown multiple testing, *Statistics & Probability Letters* 113, 38–40.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Appendix A: Description of the Data

In this appendix, we summarize the steps for the data cleaning of the Morningstar database and provide summary statistics across asset classes and investment categories.

A.1 Raw Morningstar data

Our mutual fund sample is from the Morningstar database.[14] We focus on all funds with an *Investment Type* flagged by 'Open-End Fund' or 'Exchange-Traded Fund' including non-survivors from December 1991 to December 2016. We download the following fields for each share class.

For the description of a share class we retrieve the *Name*, *ISIN*, and *Base Currency*. It is common to name a share class starting with the name of the asset manager, followed by a description of the strategy, and an ending for the share class. For example, for the equity fund 'Blackrock S&P 500 Index' there is a share class "Blackrock S&P 500 Index Institutional' for institutional and the 'Blackrock S&P 500 Index Investor A' for retail clients.

The most specific categorization within Morningstar is the *Morningstar Category*, which is derived by analyzing the underlying portfolio holdings. In total, we find 504 different groups for the retail equity and fixed income funds. The *Global Category* combines several *Morningstar Category*, and we see in total 68 groups for retail equity and fixed income funds. For example the *Global Category* category 'Europe Equity Large Cap' includes Morningstar Categories such as 'EAA Fund Europe Large-Cap Blend Equity', 'EAA Fund Europe Large-Cap Value Equity', 'EAA Fund Europe Large-Cap Growth Equity', but also 'US Fund Europe Stock' or 'Canada Fund European Equity'. Since we have within this broader categorization a higher chance to find both index and equity funds, we concentrate on the *Global Category*. The *Global Broad Category Group* further aggregates the *Global Category* into the major asset classes. Since we focus on the comparison of active and index funds, we concentrate on the *Global Broad Category Group* 'Equity' and 'Fixed Income' funds. We thereby

[14] Recent work in Kosowski et al. (2006); Fama and French (2010); Barras et al. (2010) concentrates mostly on the survivor-bias-free CRSP US Mutual Fund Database. As shown by Elton (2001), the CRSP database also suffers from a survivorship bias: the so-called omission bias. Berk and van Binsbergen (2015) find that neither the CRSP nor the Morningstar database are free of errors. Thus, we must be careful and we find the same errors as reported in this previous work.

disregard categories such as 'Allocation', 'Money Market', or 'Commodities' because we find for them only an insufficient number of index funds to make a fair comparison. We show in the Table 11 the investment categories, where we find both, index and active mutual funds.

For the return computation, we download the following fields for each fund: *Monthly Return USD*, *Monthly Gross Return USD*, and *Net Assets - share class (Monhtly) USD*. The *Monthly Return USD* includes management, administrative, and other costs that are deducted from the NAV such as the 12b-1 fee. All income and capital gains are reinvested monthly. The *Monthly Gross Return USD* is based on the *Monthly Return USD* and adds the most recent net expense ratio. The *Net Assets - share class (Monhtly) USD* is the monthly total net assets of a share class.

To distinguish between active and index funds, we make use of the *Index Fund* field. Morningstar flags funds that track a particular index based on full replication or based on a representative sampling as an index fund. Next, to filter the institutional and retail funds, we download the field *Institutional* that defines any fund as institutional fund if it either states institutional in the name of the share class, has a minimum investment above USD 100,000, or the prospectus says that it is for institutional investors only.

A.2 Data cleaning

We retrieve for every fund the monthly net return, gross return, and total net assets, all in US Dollars. We only include an observation if all three items are available. Often, and as reported in Berk and van Binsbergen (2015), we observe that net assets are reported quarterly or they are missing for a specific month. In this case, we roll the assets under the assumptions of zero net flows to increase the available data points and to avoid disconnected time series. Besides, for some institutional mutual funds, we observe zero fees because they are paid in separate contracts with the asset manager. Thus, we only include funds where the sum of the gross returns is larger than the sum of the net returns to exclude zero-fees funds. To avoid the incubation bias, we include funds only if they reach 5 million December 2016 US dollars in AUM.

We also see conversion errors, where funds assets suddenly increase by a high factor and then decrease again by a similar factor. First, we observe this behavior in emerging market currencies before 1999. Thus, we concentrate in the period before 1999 only on the developed currencies, Pound Sterling, US Dollar, Euro, Singapore Dollar, Australian Dollar, Swedish Krona, South African Rand, Swiss Franc, Japanese Yen, New Zealand Dollar, Canadian Dollar, Norwegian Krone, Danish Krone. Also, we see that for some funds, the assets change by a factor higher than 100 and decrease in the next period to the same level as before the outlier. For these cases, we smooth the net assets over time if we see that the assets change by a factor higher than 10 and we decrease in the next two periods by a factor of more than 0.5. But there are funds where this increase is verified by attaining the same fund levels in the future. Therefore, we only correct the assets, if the same level is not exceeded in the future assets.

We also delete obvious mistakes, such as when an index fund shows high fees in the past and suddenly changes to a low fee. In this case, we keep only the low fee period, since we interpret that either the fees were not correct or the fund changed from active to index.

When we build the value-weighted portfolio for the investment categories, we also remove funds that show a beta below 0.05 relative to the average return of all the funds within the same investment category. Because of the low sensitivity to the average fund, these funds do not follow a similar strategy as the rest of the group.

A.3 Aggregation of the share classes

Each line in the Morningstar data corresponds to a share class. In total, we obtain 435'453 lines of different share classes. Thus, we must aggregate the same share classes to avoid multiple tries of investment strategies of the same provider. First, we tried to use the fields *Administrator* and *Ticker of Fund's Oldest Share Class*; however, they are often missing. For this reason, we aggregate alphabetically subsequent mutual funds that are in the same *Morningstar Category* with the corresponding *Index Fund* flag and have a similar name. While Berk and van Binsbergen (2015) use the last word of the fund name for the share class, we use the ratio provided by the SequenceMatcher of the difflib library in Python, which is based on the algorithm developed by Ratcliff and Metzener (1988) and, additionally, cleans the "junk" elements. We define a similar name by a ratio above 0.8.

A.4 Summary statistics across investment categories

In Table 1 we provided the time-series averages of cross-sectional monthly attributes across asset classes. Table 11 provides a more detailed view across investment category where we find at least one index fund. For the active funds, we analyze in total 14,969 institutional and 46,300 retail funds, while we have 56,136 active and 5,133 index funds. In general, there are fewer index funds, but they show higher average total net assets (TNA) and net returns, and also lower fees and about the same average years in the database. As expected, the institutional funds charge lower fees compared to their retail counterparts.

[Table 11 about here.]

Figure 1: Total net assets per category and percentage index mutual funds

This figure presents the total net assets in US Dollar in the top graph and the percentage of index ('Index') funds in the bottom graph. We differentiate between equity ('EQ') and fixed income ('FI') funds as well as retail ('R') and institutional ('I') funds.

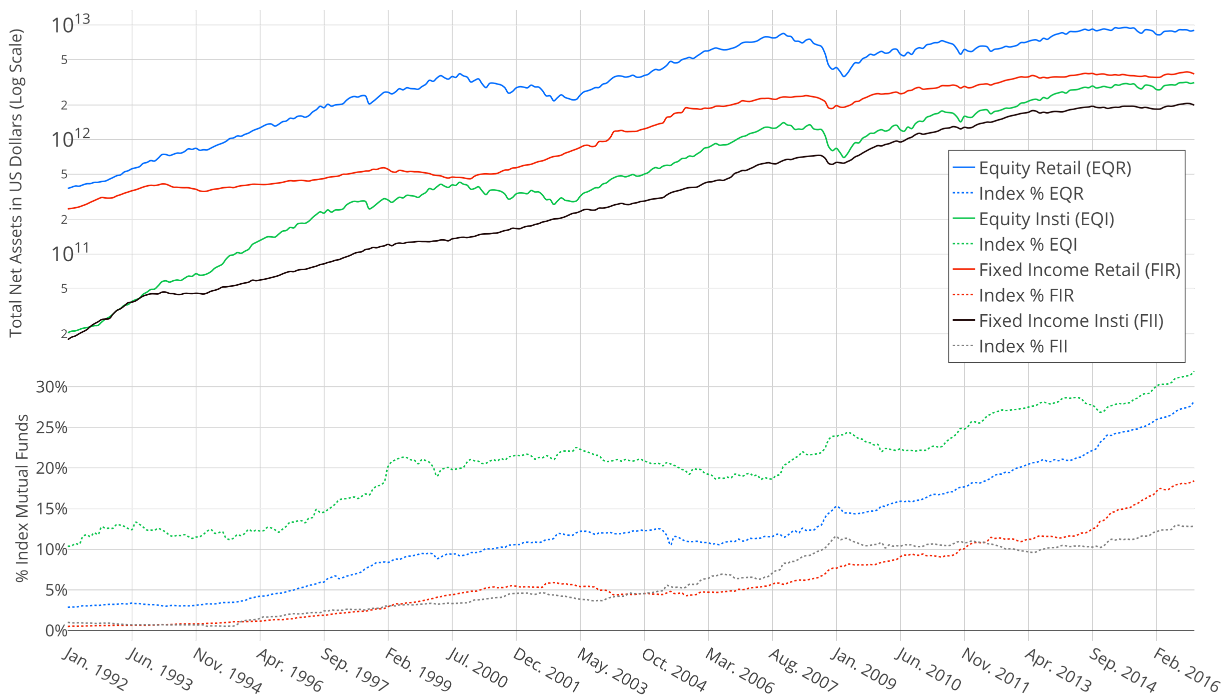

Figure 2: Optimal circular block size

This figure presents the percentage of retail mutual funds per optimal circular block length based on the method of Politis and White (2004) and Patton et al. (2009). We focus on single mutual funds' ('Single') and portfolios of mutual funds' ('Portfolio') residuals. The categories built the single equity funds ('Equity'), single fixed income funds ('Fixed Income'), the single 'US Equity Large Cap Blend' funds ('US Equity'), the value-weighted portfolios within a Morningstar investment category ('Inv. Categories'), and value-weighted portfolios of the equity funds divided by regional investment regions ('Equity Regions'). For the benchmark model we concentrate on the investable one-factor model with the value-weighted return of the index funds within the same category as the analyzed single mutual fund or portfolio of mutual funds ('Investable'), the three-factor 'FF3' and five-factor 'FF5' model with the regional factors 'market', 'size', and 'value' of Fama and French (1992) for the three-factor ('FF3'), and also 'momentum' of Jegadeesh and Titman (1993) and 'betting against beta' of Frazzini and Pedersen (2010) for the five-factor model ('FF5'). In the legend we report the category, the return type, the benchmark model, and the total number of time series within an analysis.

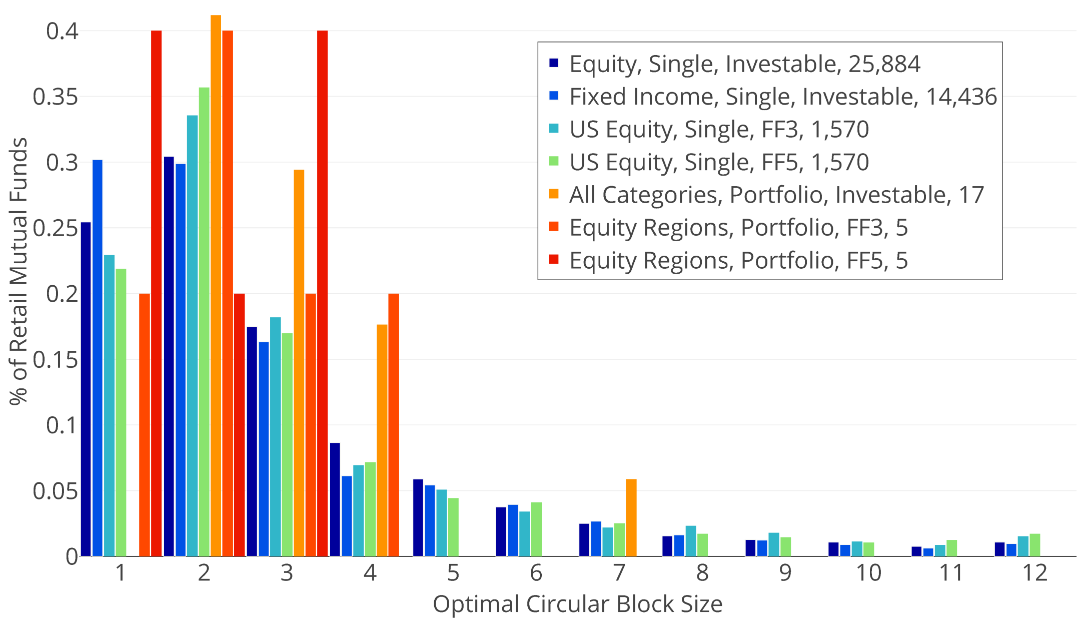

Figure 3: Equity funds: Distribution of p-values across investment regions

This figure presents the frequency of the observed p-values based on the bootstrapped significance test for the alpha. On the left we equal-weight each p-value, while on the right we value-weight the p-values by the total assets over time. The benchmark model is the five-factor model including the Fama and French (1992) factors market (MKT), size (SMB), value (HML) extended by the momentum (WML) factor of Jegadeesh and Titman (1993) and betting against beta (BAB) factor of Frazzini and Pedersen (2010). We compute the p-values for all retail and institutional funds with an assets under management (AuM) above USD 5 Mio. and more than three years of monthly net and gross return as well as assets under management data. The different investment regions are: 'US', 'Global', 'Europe', 'Japan', 'Asia ex Japan'. The number of funds per investment region (n) is provided in the legends. For each investment region, we apply the regional model for the MKT, SMB, HML, WML, and BAB factors. Data are from Kenneth French's webpage and from AQR. The negative p-values—funds with negative alphas—are highlighted in gray.

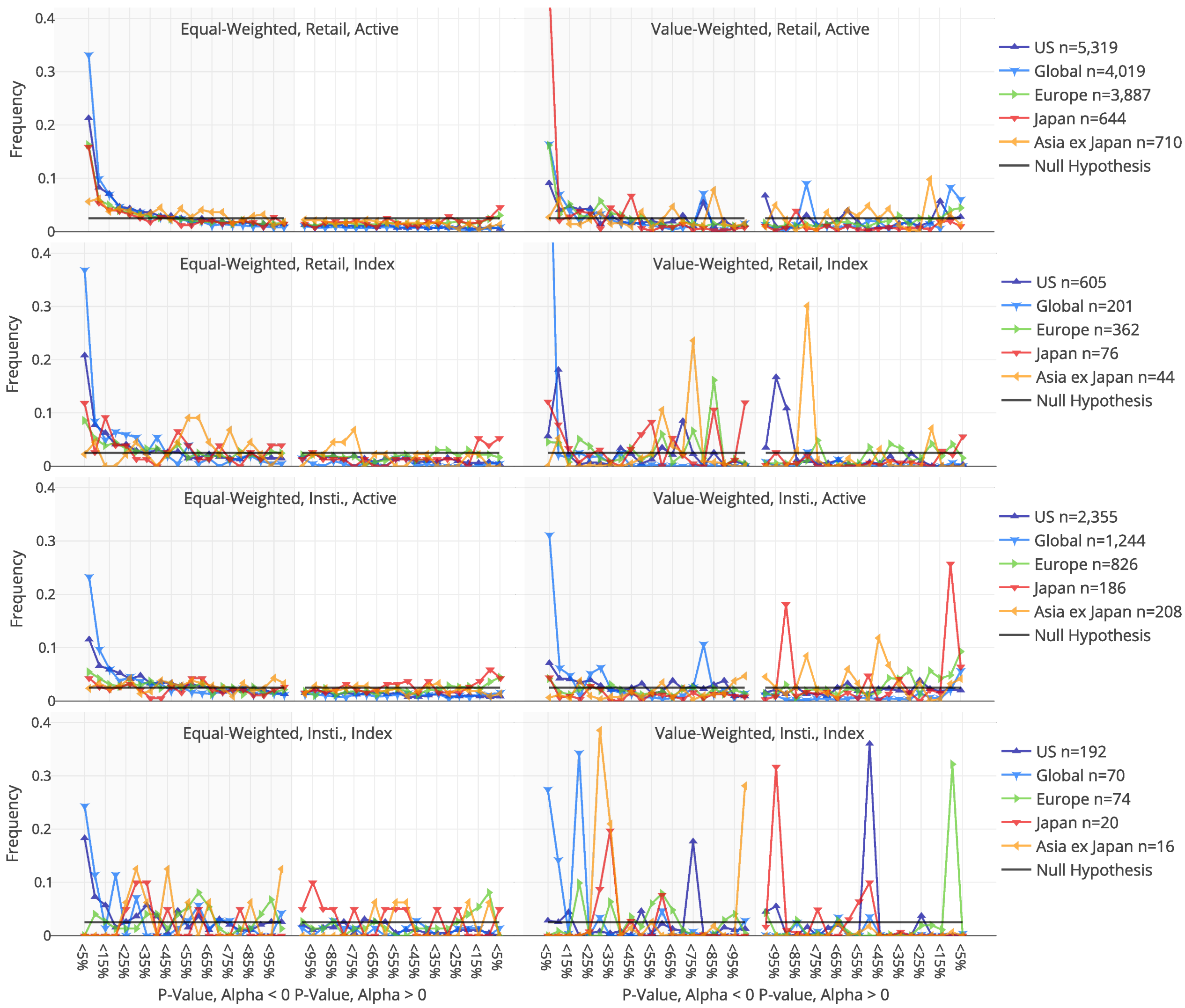

Figure 4: Value-weighted alpha of active mutual funds within investment categories

This figure presents for all investment categories as defined by the 'Global Category' of Morningstar the multiple hypothesis adjusted p-value on the y-axis and the annualized value-weighted alpha of active versus index funds on the x-axis. In the top (bottom) chart we show the analysis after (before) management fees. We form the 4 groups with the combinations retail and institutional as well as the periods 1992–2016 and 2000–2016.

Figure 5: Aggregated value-weighted alpha of active minus index

This figure presents the cumulated logarithmic returns of the active mutual funds minus the index mutual funds. We analyze both institutional and retail funds. We equal-weight ('EW') and value-weight ('VW') the Morningstar investment categories, and also regard the portfolios before ('Gross') and after ('Net') costs. We include all equity and fixed income mutual funds within Morningstar where we find the net, gross, and assets under management and at least one index fund within the same investment category. The analysis is in US dollars. Also, we show the S&P 500 index in gray bars.

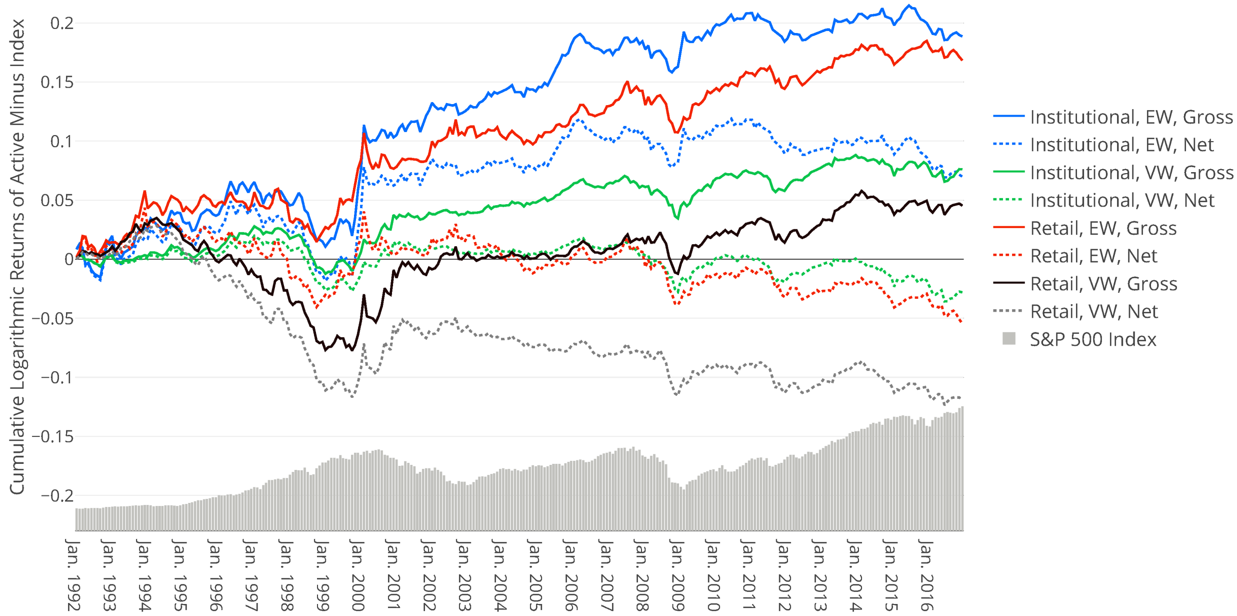

Figure 6: Retail and institutional funds with value-weighted index fund benchmark within Morningstar global category

This figure presents the frequency of the observed p-values based on the bootstrapped significance test for the alpha. The benchmark returns build the value-weighted index products within the Global Category. We compute the p-values for all retail and institutional funds with an AuM above USD 5 Mio. and more than 1 year of monthly returns. Also, we must have at least one index fund within the 'Global Category' to compute the alpha of a fund. The number of funds per investment category ('n') is provided in the legend. The negative p-values—funds with negative alphas—are highlighted in gray.

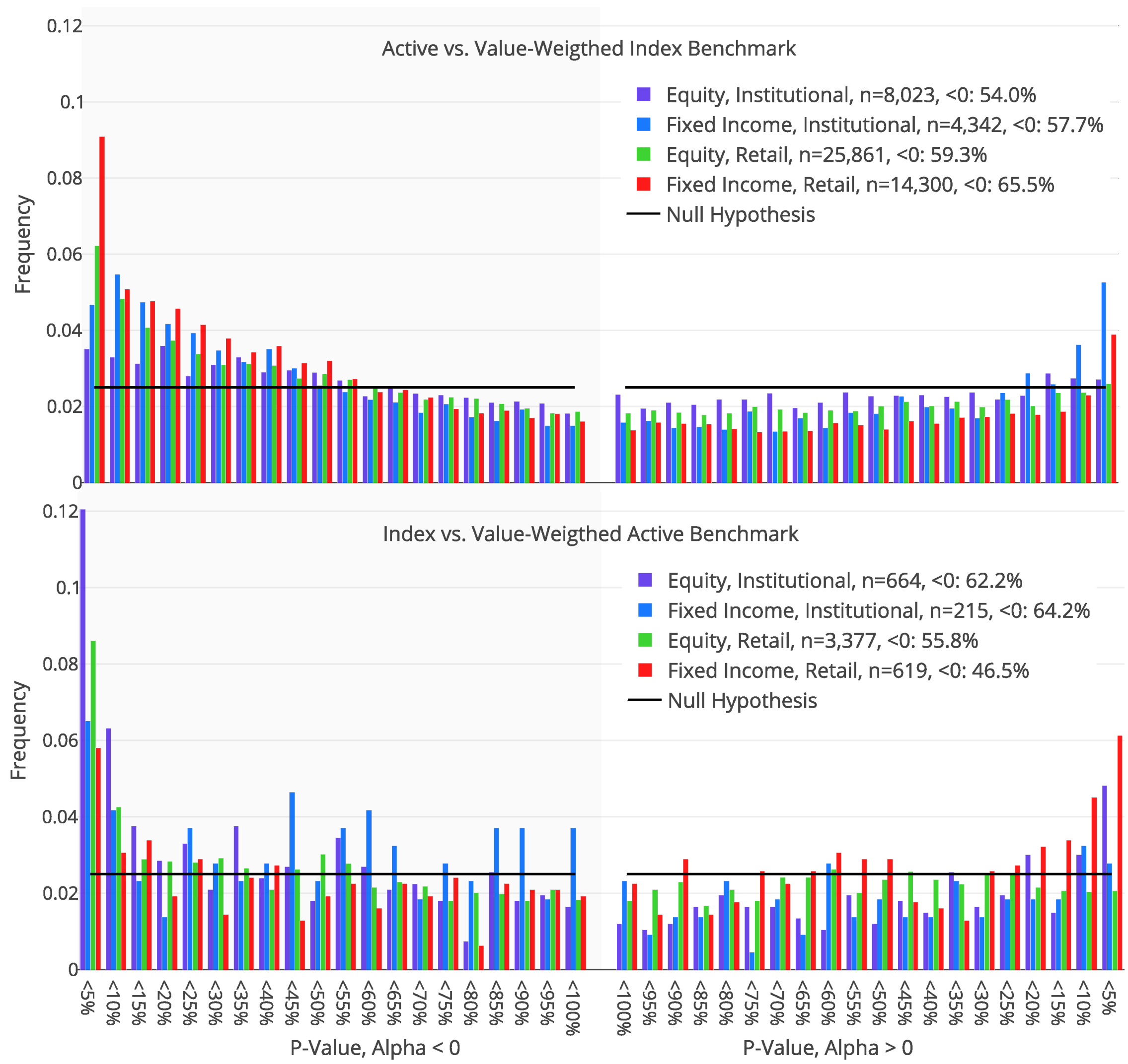

Table 1: Mutual fund database summary statistics

This table reports the average number ('Avg Number'), average total net assets in Mio. USD ('Avg TNA Mio'), average annual net return in USD ('Avg Net Ret'), average annual fee in USD ('Avg Fees ann') and the average years of a fund in the database ('Avg Years') over the time period from December 1991 to December 2016 of all available funds in the Morningstar database flagged by Open-End or Exchange-Traded funds. We only include funds within the 'Global Broad Category Group' equity ('Equity') and fixed income ('Fixed Income') for which we provide the category statistics. The average corresponds to time-series averages of cross sectional monthly attributes.

in USD		**Total Number**		**Avg Number**		**Avg TNA Mio**		**Avg Net Ret**		**Avg Fees ann**		**Avg Years**	
		Active	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*
Equity	*Inst.*	8,488	691	2,506.1	199.5	255.3	910.0	8.85%	8.95%	0.86%	0.15%	7.4	7.2
	Retail	26,741	3,551	9,147.8	950.6	453.6	732.2	8.16%	8.27%	1.18%	0.31%	8.6	6.7
Fixed	*Inst.*	5,566	224	1,440.3	57.8	333.2	663.0	4.97%	5.19%	0.54%	0.19%	6.5	6.5
Income	*Retail*	15,341	667	4,545.6	152.6	346.7	817.6	4.87%	5.20%	0.88%	0.25%	7.4	5.7

Table 2: Dependence over time

This table reports the number and percentage of mutual funds with a Fisher's p-value or Ljung-Box (LJ) p-value below 5%. The Fisher's p-value is computed as outlined in Genest and Rémillard (2004), where we use 1,000 simulations. The categories, single and portfolio approach, benchmark model, and total number of analyzed funds are the same as described in Figure 2.

Category	Equity	Fixed Income	US Equity		Inv. Categories	Equity Regions	
Single/Portfolio	Single				Portfolio		
Benchmark	Investable		FF3	FF5	Investable	FF3	FF5
Total Number	25,884	14,436	1,570	1,570	17	5	5
Fisher $<$ 5%	**4,079 (16%)**	**2,762 (19%)**	**332 (21%)**	**395 (25%)**	**6 (35%)**	**1 (20%)**	**2 (40%)**
LJ $<$ 5%	**4,626 (18%)**	**2,483 (17%)**	**335 (21%)**	**381 (24%)**	**4 (24%)**	**2 (40%)**	**2 (40%)**

Table 3: Empirical rejection probabilities: Type I errors

This table presents the empirical rejection probabilities for the nominal levels $\alpha = 10\%$, $\alpha = 5\%$, and $\alpha = 1\%$ for the standard ('Stand'), Newey-West ('NW') with a bandwith of $4 \times (T/100)^{2/9}$, HC3 ('HC3') and our bootstrapped ('Boot') significance test that evaluates the optimal block size by the method of Politis and White (2004) and the correction of Patton et al. (2009). For the DGP we sample from the realized returns with a circular bootstrap ('Boot-x') and a block size of $x = \{1, 3, 6\}$. The simulation study includes 10 US mutual funds that exhibit the entire return history from 1992 to 2016 in the Morningstar database. We sample for each fund and DGP 1,000 paths and set the alpha under the null hypothesis to the true observed alpha. We show the results for the one-factor 'CAPM', three-factor 'FF3', and five-factor 'FF5' model with the factors 'market', 'size', and 'value' of Fama and French (1992), and also the 'momentum' of Jegadeesh and Titman (1993) and 'betting against beta' factor of Frazzini and Pedersen (2010). We highlight the p-values closest to the nominal value of the test. Because the null hypothesis is true for all of the simulations, the true rejection probabilities should be equal to the nominal level of the test.

DGP	**Nominal Level**	**CAPM** *Stand*	*NW*	*HC3*	*Boot*	**FF3** *Stand*	*NW*	*HC3*	*Boot*	**FF5** *Stand*	*NW*	*HC3*	*Boot*
Boot-1	$\alpha = 0.1$	0.112	0.109	**0.099**	0.102	0.114	0.113	**0.099**	0.105	0.120	0.117	**0.096**	0.106
	$\alpha = 0.05$	0.061	0.059	**0.049**	0.053	0.062	0.060	**0.050**	0.056	0.064	0.061	0.046	**0.053**
	$\alpha = 0.01$	0.012	0.012	0.008	**0.010**	0.016	0.016	**0.012**	0.013	0.016	0.017	**0.011**	0.013
Boot-3	$\alpha = 0.1$	0.142	0.119	0.125	**0.111**	0.137	0.123	0.120	**0.106**	0.135	0.118	0.111	**0.105**
	$\alpha = 0.05$	0.084	0.065	0.070	**0.058**	0.078	0.068	0.066	**0.060**	0.076	0.062	0.058	**0.054**
	$\alpha = 0.01$	0.024	0.017	0.017	**0.013**	0.025	0.020	0.019	**0.015**	0.023	0.018	0.017	**0.013**
Boot-6	$\alpha = 0.1$	0.158	0.126	0.141	**0.115**	0.140	0.124	0.124	**0.112**	0.148	0.126	0.123	**0.114**
	$\alpha = 0.05$	0.098	0.071	0.085	**0.063**	0.078	0.071	0.066	**0.062**	0.088	0.071	0.068	**0.062**
	$\alpha = 0.01$	0.036	0.021	0.029	**0.017**	0.025	0.021	0.020	**0.018**	0.027	0.022	0.020	**0.015**

Table 4: Empirical rejection probabilities: Family wise error rates (FWER)

This table presents the empirical rejection probabilities for the nominal levels $\alpha = 10\%$, $\alpha = 5\%$, and $\alpha = 1\%$ and the multiple hypothesis framework of Romano and Wolf (2005a,b, 2016) controlling the family wise error rate based on the bootstrapped ('Boot-x') significance test with a block size of $x = \{1, 3, 6, 9, 12\}$. The DGP is a circular bootstrap ('Boot-x') with an optimal block size of $x = \{1, 3, 6\}$. The simulation study includes the 17 portfolios within the 'Inv. Categories' setting from Section 2 with the investable one-factor benchmark model that is based on the value-weighted return of index funds. We sample for each portfolio and DGP 1,000 paths and set the alpha under the null hypothesis to the true observed alpha. We highlight the p-values closest to the nominal value of the test. Because for all the simulations the null hypothesis is true, the true rejection probabilities should be equal to the nominal level of the test.

DGP	**Nominal Level**	**Boot1**	**Boot3**	**Boot6**	**Boot9**	**Boot12**
Boot-1	$\alpha = 0.1$	0.132	**0.119**	0.096	0.071	0.050
	$\alpha = 0.05$	0.066	**0.052**	0.039	0.026	0.020
	$\alpha = 0.01$	0.015	**0.009**	0.006	0.004	0.002
Boot-3	$\alpha = 0.1$	0.145	0.124	**0.098**	0.082	0.059
	$\alpha = 0.05$	0.081	0.062	**0.048**	0.040	0.030
	$\alpha = 0.01$	0.028	0.016	**0.010**	0.004	0.002
Boot-6	$\alpha = 0.1$	0.132	0.114	**0.087**	0.066	0.046
	$\alpha = 0.05$	0.073	**0.051**	0.036	0.022	0.012
	$\alpha = 0.01$	0.020	**0.008**	0.004	0.000	0.001

Table 5: Multi-factor regressions: Value-weighted retail mutual equity funds

This table shows the regression results of the one-factor, three-factor, and five-factor model including the Fama and French (1992) factors market (MKT), size (SMB), value (HML) extended by the momentum (WML) factor of Jegadeesh and Titman (1993) and betting against beta (BAB) factor of Frazzini and Pedersen (2010). The analysis is based on value-weighted returns of retail active ('Active) and index ('Index') funds. Coefficient estimates are multiplied by 100 and HC3 Standard errors in parentheses: * p<0.1, ** p<0.05, ***p<0.01.

		US		**Global**		**Europe**		**Japan**		**Asia ex-Japan**	
		Active	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*
1-factor	*Alpha*	-0.08***	-0.01	-0.03	-0.17***	-0.15**	-0.06	-0.10	-0.10	-0.26*	-0.16
		(0.03)	(0.03)	(0.06)	(0.06)	(0.06)	(0.04)	(0.08)	(0.09)	(0.14)	(0.15)
	MKT	99.94***	98.82***	100.31***	103.38***	102.55***	100.17***	102.22***	97.41***	97.17***	94.97***
		(0.9)	(0.77)	(1.56)	(1.34)	(1.21)	(0.88)	(1.7)	(2.11)	(2.5)	(3.83)
3-factor	*Alpha*	-0.09***	-0.02	-0.05	-0.20***	-0.13**	-0.06	-0.05	-0.10	-0.34**	-0.16
		(0.03)	(0.02)	(0.06)	(0.06)	(0.06)	(0.04)	(0.08)	(0.09)	(0.13)	(0.16)
	MKT	98.69***	100.56***	100.55***	104.03***	103.56***	98.73***	100.77***	98.28***	96.49***	95.24***
		(0.86)	(0.57)	(1.41)	(1.3)	(1.23)	(0.79)	(1.77)	(2.09)	(2.53)	(3.78)
	SMB	7.20***	-7.25***	16.55***	0.76	5.01	-14.11***	-3.41	-12.54***	1.93	-4.39
		(1.21)	(1.29)	(3.58)	(3.37)	(3.25)	(1.85)	(3.18)	(3.06)	(6.39)	(6.33)
	HML	-1.53	5.97***	-2.15	7.77**	-6.38**	4.03**	-8.87***	9.08***	12.94***	-1.85
		(1.28)	(0.96)	(4.02)	(3.08)	(2.73)	(1.64)	(3.29)	(3.26)	(4.64)	(5.77)
5-factor	*Alpha*	-0.11***	-0.02	-0.06	-0.25***	-0.10	-0.04	-0.02	-0.05	-0.30*	-0.07
		(0.03)	(0.02)	(0.06)	(0.06)	(0.06)	(0.04)	(0.07)	(0.08)	(0.16)	(0.21)
	MKT	99.41***	100.16***	101.34***	104.80***	103.70***	98.42***	100.01***	97.03***	95.96***	94.77***
		(0.93)	(0.58)	(1.75)	(1.41)	(1.37)	(0.82)	(1.79)	(2.07)	(2.65)	(4.08)
	SMB	7.32***	-6.91***	16.71***	-1.00	10.05***	-13.91***	-2.42	-9.94***	2.44	-2.95
		(1.23)	(1.15)	(3.3)	(3.51)	(3.31)	(1.98)	(3.08)	(2.94)	(6.74)	(6.74)
	HML	-2.44	4.68***	2.79	6.19*	-4.65*	3.50**	-7.90***	4.73*	11.56**	-2.77
		(1.51)	(0.83)	(3.39)	(3.56)	(2.65)	(1.69)	(2.93)	(2.87)	(5.46)	(6.46)
	WML	0.37	-1.95***	5.28**	1.12	1.34	-1.37	5.33**	-6.72***	-2.8	-2.26
		(1.12)	(0.63)	(2.41)	(2.1)	(1.66)	(1.00)	(2.32)	(2.28)	(4.29)	(4.67)
	BAB	2.45*	1.35**	-4.82	4.51	-6.76***	-0.12	-7.65***	-4.30**	-0.47	-6.89
		(1.35)	(0.63)	(3.27)	(3.24)	(2.25)	(1.45)	(2.16)	(1.99)	(5.71)	(7.62)

Table 6: Multi-factor regressions: Value-weighted institutional mutual equity funds

This table shows the regression results of the one-factor, three-factor and five-factor model including the Fama and French (1992) factors market (MKT), size (SMB), value (HML) extended by the momentum (WML) factor of Jegadeesh and Titman (1993) and betting against beta (BAB) factor of Frazzini and Pedersen (2010). The analysis is based on value-weighted returns of institutional active ('Active) and index ('Index') funds. Coefficient estimates are multiplied by 100 and HC3 Standard errors in parentheses: * p<.1, ** p<.05, ***p<.01.

		US		**Global**		**Europe**		**Japan**		**Asia ex-Japan**	
		Active	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*
1-factor	*Alpha*	-0.08**	0.00	-0.07	-0.20***	0.07	-0.02	0.02	-0.01	-0.02	-0.07
		(0.04)	(0.03)	(0.06)	(0.07)	(0.06)	(0.06)	(0.08)	(0.07)	(0.12)	(0.07)
	MKT-RF	101.73***	97.23***	101.06***	102.69***	95.45***	95.57***	104.18***	98.66***	95.03***	96.83***
		(1.05)	(0.93)	(1.5)	(1.46)	(1.47)	(1.2)	(1.86)	(1.73)	(1.85)	(1.81)
3-factor	*Alpha*	-0.11***	-0.11***	-0.13**	-0.25***	0.03	0.02	-0.02	0.00	-0.08	-0.03
		(0.03)	(0.02)	(0.06)	(0.07)	(0.06)	(0.05)	(0.08)	(0.07)	(0.13)	(0.06)
	MKT-RF	99.77***	99.81***	102.39***	103.74***	96.37***	94.37***	105.40***	99.46***	95.93***	98.15***
		(0.83)	(0.62)	(1.25)	(1.34)	(1.35)	(0.93)	(1.73)	(1.27)	(1.94)	(1.04)
	SMB	14.76***	-12.41***	16.81***	0.27	14.32***	-18.88***	-3.81	-15.19***	-2.20	-21.20***
		(1.38)	(1.2)	(3.98)	(4.31)	(4.33)	(2.11)	(3.31)	(2.76)	(4.79)	(2.45)
	HML	2.37*	6.58***	10.42***	11.52***	-0.72	0.72	8.78**	8.09***	12.25**	-10.31***
		(1.38)	(1.02)	(3.97)	(3.6)	(3.07)	(2.18)	(4.16)	(2.45)	(5.53)	(3)
5-factor	*Alpha*	-0.13***	-0.01	-0.19***	-0.29***	0.02	0.02	0.02	0.00	0.05	-0.07
		(0.03)	(0.02)	(0.06)	(0.07)	(0.07)	(0.05)	(0.07)	(0.07)	(0.15)	(0.07)
	MKT-RF	100.20***	99.37***	104.04***	104.84***	97.29***	94.19***	102.64***	99.15***	95.23***	98.32***
		(0.87)	(0.64)	(1.37)	(1.49)	(1.32)	(1.02)	(1.91)	(1.17)	(1.95)	(1.08)
	SMB	15.00***	-12.01***	14.45***	-1.44	16.87***	-20.10***	1.75	-13.42***	1.17	-22.15***
		(1.37)	(1.09)	(4.07)	(4.41)	(4.32)	(2.48)	(3.36)	(2.36)	(5.05)	(2.7)
	HML	1.06	5.04***	10.63***	11.08***	2.17	-0.18	5.81	7.34***	12.23**	-10.38***
		(1.54)	(0.88)	(3.69)	(3.61)	(3.14)	(2.27)	(3.58)	(2.56)	(5.51)	(3.02)
	WML	-0.50	-2.24***	3.94	2.23	4.67**	-1.18	-6.54**	-4.56**	-0.22	-0.24
		(0.93)	(0.61)	(2.39)	(2.25)	(1.85)	(1.00)	(2.81)	(1.99)	(4.12)	(2.58)
	BAB	2.64**	1.68**	4.16	3.54	-4.89**	2.07	-9.10***	-0.1	-9.98*	2.82
		(1.12)	(0.69)	(3.36)	(3.3)	(2.32)	(1.64)	(2.26)	(1.93)	(5.5)	(2.59)

Table 7: Value-weighted alpha across investment regions

This table shows the multiple hypothesis analysis of the value-weighted alpha of equity funds across the investment regions: US, Global, Europe, Japan, and Asia ex-Japan. We show for each of the categories the gross and net return alpha for active, and index funds, and the for multiple tries adjusted p-value of Romano and Wolf (2016) based on the robust alpha test with a block size of six. Regarding the benchmark model, we focus on one-, three-, and five-factor models with the Fama and French (1992) factors market (MKT), size (SMB), value (HML) extended by the momentum (WML) factor of Jegadeesh and Titman (1993) and betting against beta (BAB) factor of Frazzini and Pedersen (2010). We highlight significant adjusted p-values below the 10% significance level in bold. The numbers are based on monthly log-returns and shown in percentage numbers. Furthermore, we distinguish between institutional and retail funds. The analysis starts in February 1998 and it ends in December 2016. For the regions and categories without a value, we miss the entire data history.

		[in %]		**US**		**Global**		**Europe**		**Japan**		**Asia ex-Japan**	
				Active	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*
AFTER COSTS	**1-factor**	**Insti.**	*Alpha*	-0.07	-0.01	-0.07	**-0.16**	0.07	-0.01	-	-	-	-
			p-adj	32.47	94.10	45.13	**2.22**	45.13	94.10	-	-	-	-
		Retail	*Alpha*	-0.07	-0.02	-0.08	**-0.18**	**-0.14**	**-0.09**	-0.12	-0.11	-0.08	-0.12
			p-adj	21.28	71.67	33.35	**3.64**	**7.68**	**8.76**	21.28	46.69	65.77	71.67
	3-factors	**Insti.**	*Alpha*	**-0.10**	0.00	**-0.14**	**-0.20**	0.04	0.03	-	-	-	-
			p-adj	**2.04**	87.62	**7.40**	**0.68**	54.25	76.82	-	-	-	-
		Retail	*Alpha*	-0.08	-0.02	-0.10	**-0.21**	-0.13	-0.07	-0.07	-0.11	-0.17	-0.13
			p-adj	17.48	59.11	35.71	**1.70**	13.44	12.26	44.19	47.09	44.19	59.11
	5-factors	**Insti.**	*Alpha*	**-0.12**	-0.01	**-0.19**	**-0.24**	0.04	0.02	-	-	-	-
			p-adj	**1.16**	85.90	**1.64**	**0.40**	60.83	85.90	-	-	-	-
		Retail	*Alpha*	-0.09	-0.02	-0.12	**-0.25**	-0.11	-0.06	-0.04	-0.07	-0.08	-0.10
			p-adj	12.28	70.17	23.76	**0.38**	17.16	30.65	70.01	70.17	70.01	70.17
BEFORE COSTS	**1-factor**	**Insti.**	*Alpha*	-0.01	0.00	0.00	**-0.13**	**0.15**	0.00	-	-	-	-
			p-adj	98.30	99.98	99.62	**7.14**	**6.70**	99.98	-	-	-	-
		Retail	*Alpha*	0.01	0.00	0.02	-0.15	0.00	-0.05	0.02	-0.05	0.06	-0.06
			p-adj	99.68	95.88	99.68	10.22	99.68	61.41	99.68	89.74	99.68	92.58
	3-factors	**Insti.**	*Alpha*	-0.04	0.01	-0.07	**-0.18**	0.12	0.04	-	-	-	-
			p-adj	52.49	76.58	52.49	**2.24**	24.82	54.05	-	-	-	-
		Retail	*Alpha*	0.01	0.00	0.00	**-0.18**	0.01	-0.02	0.06	-0.05	-0.03	-0.06
			p-adj	99.92	95.96	99.92	**3.82**	99.92	86.92	81.72	88.50	99.92	90.66
	5-factors	**Insti.**	*Alpha*	-0.05	0.00	-0.12	**-0.21**	0.11	0.03	-	-	-	-
			p-adj	23.66	87.28	23.66	**0.78**	23.66	64.91	-	-	-	-
		Retail	*Alpha*	-0.01	0.00	-0.02	**-0.22**	0.03	-0.01	0.09	-0.02	0.06	-0.04
			p-adj	97.56	99.34	97.56	**0.80**	97.56	98.36	46.17	99.34	97.56	99.34

Table 8: Single mutual equity funds: Zero alpha, skilled, and unskilled

This table shows the proportion of zero alpha, skilled and unskilled funds across the investment regions 'US', 'Global', 'Europe', 'Japan', and 'Asia ex-Japan'. The results are based on the method of Barras et al. (2010), while we apply the robust alpha test presented in this work to compute the single mutual funds' p-values. We show the results for the five-factors benchmark model including the regional model of the Fama & French for MKT, SMB, HML and WML homepage and AQR homepage for BAB.

		5-factors	**US**	**Global**	**Europe**	**Japan**	**Asia ex-Japan**	**Average**
Retail	**Active**	*Zero alpha*	55.1%	39.6%	66.2%	67.9%	83.0%	*62.3%*
		Skilled	0.0%	0.0%	3.0%	5.7%	0.0%	*1.7%*
		Unskilled	44.9%	60.4%	30.8%	26.4%	17.0%	*35.9%*
	Index	*Zero alpha*	61.9%	30.1%	76.5%	73.7%	100.0%	*68.4%*
		Skilled	0.0%	0.0%	3.6%	5.9%	0.0%	*1.9%*
		Unskilled	38.1%	69.9%	19.9%	20.4%	0.0%	*29.7%*
Institutional	**Active**	*Zero alpha*	69.3%	53.5%	78.5%	88.2%	97.4%	*77.4%*
		Skilled	0.0%	0.0%	8.2%	9.4%	0.0%	*3.5%*
		Unskilled	30.7%	46.5%	13.3%	2.4%	2.6%	*19.1%*
	Index	*Zero alpha*	66.9%	55.7%	91.9%	92.5%	90.6%	*79.5%*
		Skilled	0.0%	0.0%	6.8%	0.0%	0.0%	*1.4%*
		Unskilled	33.1%	44.3%	1.4%	7.5%	9.4%	*19.1%*

Table 9: Active and index investing: Zero alpha, skilled, and unskilled

This table shows the proportion of zero alpha, skilled and unskilled funds for all retail and institutional funds with an AuM above USD 5 Mio. and more than one year of monthly returns. We distinguish between equity and fixed income mutual funds as well as institutional and retail investors. The results are based on the method developed by Barras et al. (2010), while we apply the robust alpha test presented in this work to compute the single mutual funds' p-values. We show the results when we compare the single active mutual funds against the value-weighted performance of index mutual funds within the same investment category and vice versa.

		Institutional		**Retail**		**Average**
		Equity	*Fixed Income*	*Equity*	*Fixed Income*	
Active	*Zero alpha*	89.3%	71.2%	82.5%	71.9%	*78.7%*
	Skilled	2.4%	9.0%	1.8%	3.2%	*4.1%*
	Unskilled	8.3%	19.8%	15.7%	24.9%	*17.2%*
Index	*Zero alpha*	70.6%	91.4%	85.6%	84.0%	*82.9%*
	Skilled	6.7%	0.0%	1.6%	8.9%	*4.3%*
	Unskilled	22.7%	8.6%	12.8%	7.1%	*12.8%*

Table 10: One year persistence after fees

This table reports the yearly performance difference, annualized alpha, bootstrapped single p-value and the beta for the value-weighted performance of active mutual funds against the one-factor benchmark model with the value-weighted performance of index funds ('Index') in the top, and the value-weighted performance of active mutual funds ('All') in the bottom of the table. Each row from 90% to 10% corresponds to the value weighted portfolio filtered by the x% best mutual funds of the past year. We rebalance the portfolios that exclude a certain percentage of funds every year in December. The row 'All' includes all active funds.

	1993 - 2016		**All**	**90%**	**80%**	**70%**	**60%**	**50%**	**40%**	**30%**	**20%**	**10%**
Index	*perf. diff. p.a.*	*Insti.*	0.08%	0.19%	0.22%	0.27%	0.40%	0.56%	0.68%	0.77%	0.80%	1.20%
		Retail	-0.57%	-0.46%	-0.39%	-0.32%	-0.26%	-0.25%	-0.21%	-0.15%	-0.03%	0.23%
	alpha p.a.	*Insti.*	-0.21%	-0.11%	-0.08%	-0.03%	-0.02%	-0.03%	0.02%	0.05%	0.03%	0.20%
		Retail	-0.55%	-0.38%	-0.30%	-0.23%	-0.19%	-0.20%	-0.19%	-0.23%	-0.22%	-0.06%
	p-value	*Insti.*	0.44	0.70	0.77	0.92	0.95	0.94	0.96	0.92	0.96	0.79
		Retail	0.18	0.38	0.49	0.59	0.69	0.68	0.73	0.69	0.76	0.95
	beta	*Insti.*	1.01	1.01	1.01	1.01	1.02	1.02	1.03	1.03	1.03	1.04
		Retail	1.00	0.99	0.99	0.99	0.99	1.00	1.00	1.01	1.03	1.04
All	*perf. diff. p.a.*	*Insti.*		0.10%	0.14%	0.18%	0.21%	0.23%	0.31%	0.37%	0.36%	0.60%
		Retail		0.11%	0.18%	0.26%	0.32%	0.32%	0.36%	0.43%	0.55%	0.81%
	alpha p.a.	*Insti.*		0.10%	0.13%	0.18%	0.19%	0.18%	0.23%	0.26%	0.23%	0.39%
		Retail		0.16%	0.24%	0.31%	0.34%	0.33%	0.33%	0.29%	0.30%	0.41%
	p-value	*Insti.*		0.06	0.19	0.29	0.4	0.55	0.55	0.61	0.69	0.6
		Retail		0.03	0.03	0.02	0.11	0.33	0.36	0.51	0.63	0.61
	beta	*Insti.*		1.00	1.00	1.00	1.00	1.01	1.01	1.02	1.02	1.03
		Retail		0.99	0.99	0.99	1.00	1.00	1.00	1.02	1.03	1.06

Table 11: Mutual fund investment categories summary statistics

This table reports the total number ('Total Number'), average number ('Avg Number'), average total net assets in USD ('Avg TNA Mio'), average annual net return in USD ('Avg Net Ret'), the average years of a fund in the database ('Avg Years'), and the first appearance of an index fund over the time period from December 1991 to December 2016 of all available investment categories ('Global Category') within the Morningstar database. We only include funds within the 'Global Broad Category Group' equity ('Equity') or fixed income ('Fixed Income') that are flagged by Open-End or Exchange-Traded funds. The average corresponds to time-series averages of cross sectional monthly attributes.

in USD		**Avg Number**		**Avg TNA Mio**		**Avg Net Ret**		**Avg Years**		**First Index**
		Active	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*	**Fund**
Mexico Fixed Income	*Retail*	182.3	2.0	233.7	137.4	-1.89%	1.28%	5.8	7.5	May 09
Global Equity Large Cap	*Inst.*	363.2	25.8	435.8	685.3	6.73%	6.17%	7.1	6.0	Jul 94
	Retail	1,093.9	57.6	466.4	603.6	6.22%	5.99%	7.5	5.6	Jan 92
Mexico Equity	*Inst.*	22.7	6.2	80.0	26.4	8.93%	9.33%	6.8	6.7	Feb 04
	Retail	26.0	10.6	70.7	297.7	7.01%	9.55%	7.0	8.5	Feb 04
Global Fixed Income	*Inst.*	329.7	21.6	295.9	241.4	4.36%	3.85%	5.2	5.2	Apr 03
	Retail	1,104.1	33.9	201.9	242.3	2.63%	2.68%	5.5	4.4	Apr 05
Europe Equity Large Cap	*Inst.*	211.9	24.5	126.2	427.5	6.97%	6.34%	6.0	6.0	Jan 98
	Retail	812.6	73.8	237.7	425.9	6.71%	7.33%	7.9	6.6	Jan 92
Euro Fixed Income	*Inst.*	349.7	19.1	298.7	638.7	2.43%	3.40%	6.0	6.6	Apr 04
	Retail	1,103.6	23.7	354.7	241.2	3.83%	5.71%	8.3	5.6	Jan 92
US Equity Small Cap	*Inst.*	234.2	13.3	252.6	459.8	11.88%	11.12%	9.6	9.5	Oct 92
	Retail	456.0	43.0	280.5	580.4	10.78%	10.85%	11.0	10.2	Jan 92
Global Equity	*Inst.*	165.0	2.1	251.2	236.9	3.86%	0.04%	5.5	5.1	Jun 07
	Retail	382.6	19.5	706.9	224.1	6.68%	5.65%	9.0	5.7	Aug 95
High Yield Fixed Income	*Retail*	954.2	16.7	307.3	197.2	4.57%	3.24%	5.1	3.6	Dec 07
Other Fixed Income	*Inst.*	214.3	4.2	237.7	78.5	2.16%	3.09%	4.3	4.1	Apr 05
	Retail	651.8	10.8	202.0	59.6	2.12%	2.39%	4.6	3.3	Apr 05
US Equity Mid Cap	*Inst.*	188.1	10.8	254.7	848.5	10.89%	11.92%	9.0	7.5	Dec 92
	Retail	487.2	39.3	450.5	778.7	10.02%	11.00%	10.5	7.5	Jan 92
Other Europe Equity	*Inst.*	78.7	10.5	95.9	665.1	7.49%	6.10%	6.5	7.0	Dec 98
	Retail	547.9	68.6	281.7	209.3	9.03%	8.24%	9.1	8.0	Jan 92
Financials Sector Equity	*Inst.*	22.2	2.4	25.7	46.4	5.59%	3.74%	7.3	10.1	Mar 04
	Retail	113.0	32.8	138.8	114.9	5.29%	5.60%	8.7	6.9	Feb 01
Africa Fixed Income	*Inst.*	73.7	1.6	131.6	11.0	3.69%	3.65%	5.6	6.6	Jan 09
	Retail	90.5	1.7	90.7	22.5	5.02%	5.81%	6.3	3.4	Nov 08
Islamic Equity	*Retail*	70.5	9.5	76.4	31.0	3.65%	3.29%	6.8	7.3	Feb 07
Other Sector Equity	*Inst.*	69.1	1.6	56.2	24.4	4.08%	4.84%	5.3	4.4	Mar 07
	Retail	262.4	40.7	124.0	136.6	4.46%	2.02%	6.2	5.1	Nov 05

Table 9 (continued)

in USD		Avg Number		Avg TNA Mio		Avg Net Ret		Avg Years		First Index
		Active	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*	**Fund**
Africa Equity	Inst.	106.9	6.8	70.7	12.3	0.66%	2.53%	5.5	4.4	Nov 07
	Retail	145.6	17.6	94.6	71.4	12.94%	13.52%	7.3	6.5	Apr 03
Technology Sector Equity	Inst.	38.9	5.3	57.3	29.2	8.94%	7.29%	6.5	9.6	Apr 04
	Retail	165.8	20.7	295.4	182.8	12.51%	11.56%	9.3	8.0	Jan 92
Energy Sector Equity	Inst.	40.5	1.9	103.9	145.9	6.84%	9.06%	5.7	7.8	Nov 04
	Retail	120.9	22.2	219.4	249.5	6.83%	7.39%	8.0	6.7	Jul 00
US Equity Large Cap Growth	Inst.	255.3	6.2	378.6	757.8	6.57%	5.32%	8.8	7.2	Jun 98
	Retail	551.4	21.8	1'229.6	809.9	8.65%	9.82%	11.0	8.1	Dec 92
US Equity Large Cap Value	Inst.	194.3	4.8	415.7	863.6	7.10%	7.16%	8.3	8.0	Aug 98
	Retail	381.8	21.8	1'006.4	758.0	8.62%	10.15%	10.6	6.9	Dec 92
US Fixed Income	Inst.	403.3	16.5	573.4	1'387.8	4.92%	5.35%	9.7	8.6	Jan 92
	Retail	843.9	50.5	540.4	1'540.5	4.59%	5.26%	10.6	7.6	Jan 92
Other Europe Fixed Income	Inst.	81.6	13.3	233.5	1'000.8	5.85%	6.63%	6.8	7.7	Nov 01
	Retail	263.2	7.7	327.7	185.2	4.30%	5.32%	8.2	6.9	Mar 98
US Equity Large Cap Blend	Inst.	233.1	54.6	289.9	2'019.1	8.68%	9.27%	8.4	9.8	Jan 92
	Retail	692.4	146.2	629.6	1'912.6	7.84%	8.98%	9.1	8.7	Jan 92
Asia Equity	Inst.	23.4	1.2	86.3	750.2	3.19%	3.91%	6.1	10.2	Jun 00
	Retail	111.5	6.2	142.0	758.8	4.66%	2.69%	9.1	6.7	Jan 92
Real Estate Sector Equity	Inst.	163.0	12.9	221.8	359.3	9.28%	8.21%	7.1	6.0	Feb 04
	Retail	268.8	26.8	162.3	628.0	9.86%	10.51%	8.2	6.3	Jun 96
Inflation Linked	Inst.	108.6	8.2	303.9	222.0	3.07%	3.46%	6.5	4.8	Feb 04
	Retail	160.9	15.8	284.4	391.6	4.03%	4.60%	7.8	6.2	Dec 98
Emerging Markets Fixed Income	Inst.	457.1	2.5	261.1	205.3	-0.82%	4.03%	2.8	1.8	Jun 13
	Retail	457.5	9.2	174.8	500.0	5.40%	7.32%	5.5	3.8	Mar 04
Emerging Markets Equity	Inst.	231.6	10.6	372.5	495.3	8.64%	8.56%	6.3	4.7	Jul 00
	Retail	417.3	29.5	214.4	427.7	7.69%	7.13%	7.5	5.1	May 92
Asia ex-Japan Equity	Inst.	115.1	10.1	155.3	169.1	12.09%	11.28%	6.0	7.3	Apr 03
	Retail	282.9	16.4	176.6	96.0	5.82%	7.09%	7.3	6.3	Nov 94
Greater China Equity	Inst.	94.1	2.3	71.7	43.5	10.84%	11.45%	5.2	4.5	Apr 09
	Retail	234.5	94.5	188.5	1'084.6	10.27%	10.73%	5.7	4.1	Jan 01
Japan Equity	Inst.	85.5	10.8	129.3	253.5	1.78%	2.31%	5.7	5.4	May 00
	Retail	300.9	37.6	262.0	301.1	3.84%	3.28%	6.9	5.9	Feb 98
UK Equity Large Cap	Inst.	59.1	8.1	324.4	542.7	5.32%	4.55%	6.1	6.4	Jan 06
	Retail	170.6	36.8	423.9	445.4	3.86%	3.52%	6.3	7.3	Nov 99

Table 9 (continued)

in USD		Avg Number *Active*	*Index*	Avg TNA Mio *Active*	*Index*	Avg Net Ret *Active*	*Index*	Avg Years *Active*	*Index*	First Index Fund
Global Equity Mid/Small Cap	Inst.	103.3	3.3	344.8	223.4	11.47%	13.73%	4.9	4.3	May 09
	Retail	288.3	9.8	230.3	223.5	4.34%	6.45%	6.4	5.2	Jul 06
Asia Fixed Income	Inst.	30.6	2.5	85.6	292.8	3.89%	1.81%	4.6	9.9	Apr 05
	Retail	268.6	6.8	163.8	110.8	5.72%	4.50%	4.9	3.7	Jan 06
Cons. Goods & Serv. Sect. Eq.	Inst.	16.5	9.8	85.4	23.7	8.22%	7.30%	5.3	9.7	Mar 04
	Retail	97.0	32.0	98.2	210.0	7.41%	6.28%	7.1	7.4	Jul 00
Sterling Fixed Income	Inst.	51.7	6.1	275.6	98.6	2.63%	2.04%	5.5	7.1	Apr 05
	Retail	169.7	15.3	447.3	457.7	1.72%	2.18%	6.2	5.8	Apr 05
Europe Equity Mid/Small Cap	Inst.	130.1	1.3	103.5	64.1	7.95%	6.91%	2.7	2.7	Dec 12
	Retail	350.2	7.8	133.1	49.7	7.57%	8.75%	7.6	6.8	Jun 01
Latin America Equity	Inst.	31.3	1.0	127.5	4.6	1.41%	-0.42%	4.1	7.6	Aug 07
	Retail	86.1	10.3	140.3	764.1	8.07%	10.33%	8.1	6.3	Aug 00
Natural Resources Sector Equity	Inst.	37.5	2.0	117.1	45.8	7.65%	8.22%	6.8	8.5	Mar 04
	Retail	107.5	14.5	167.2	162.9	7.63%	6.51%	8.9	6.6	Apr 94
Brazil Equity	Retail	51.4	6.2	54.5	21.9	-0.17%	3.29%	6.5	4.9	Aug 07
India Equity	Inst.	33.7	1.0	89.1	26.0	16.92%	17.15%	1.3	2.4	Sep 08
	Retail	169.7	9.4	169.3	233.7	8.36%	7.66%	7.5	5.2	Jan 07
Utilities Sector Equity	Inst.	15.0	1.6	64.8	143.5	8.51%	8.42%	8.6	9.9	May 04
	Retail	32.6	8.1	644.6	308.8	7.38%	8.80%	11.3	8.1	Jan 92
Healthcare Sector Equity	Inst.	32.8	4.5	77.9	62.1	9.76%	6.76%	6.5	9.7	Mar 04
	Retail	173.5	22.2	415.9	342.7	7.25%	7.11%	8.3	6.1	Jun 00
UK Equity Mid/Small Cap	Inst.	32.5	1.0	151.4	61.7	0.89%	-0.62%	1.2	1.9	Feb 15
	Retail	143.2	4.6	328.5	112.0	6.56%	6.65%	6.5	5.7	Jan 06
Communications Sector Equity	Inst.	4.6	4.9	9.5	8.9	7.25%	6.67%	6.0	9.5	Apr 05
	Retail	39.5	13.7	148.8	73.3	5.73%	8.87%	7.9	9.1	Oct 01
Korea Equity	Inst.	83.5	5.3	33.9	29.9	6.93%	6.05%	4.5	4.4	Mar 07
	Retail	280.0	41.0	96.9	57.1	13.67%	13.23%	8.9	5.7	May 01
Asia Pacific Fixed Income	Inst.	21.6	1.1	108.7	22.3	3.05%	3.75%	3.7	5.5	Sep 06
	Retail	120.9	9.6	39.6	130.1	2.60%	3.34%	7.2	5.9	May 05
Thailand Equity	Retail	117.5	8.8	47.0	49.0	17.88%	17.17%	10.9	9.3	Jan 01
Other Asia Equity	Retail	78.6	3.5	89.8	30.7	1.46%	1.06%	5.2	6.3	Jan 08
Precious Metals Sector Equity	Retail	76.8	5.9	196.1	376.0	12.31%	10.76%	9.6	6.1	Jan 92
Canadian Equity Large Cap	Inst.	4.7	2.8	34.0	216.2	5.03%	11.24%	4.9	6.5	Apr 03
	Retail	159.2	13.3	368.8	379.4	8.61%	8.79%	11.1	6.6	Jan 92

Table 9 (continued)

in USD		Avg Number		Avg TNA Mio		Avg Net Ret		Avg Years		First Index
		Active	*Index*	*Active*	*Index*	*Active*	*Index*	*Active*	*Index*	**Fund**
Thailand Fixed Income	Retail	82.8	1.0	126.6	165.0	3.60%	6.25%	7.2	10.8	Mar 06
South American Equity	Retail	35.4	1.2	39.0	16.9	4.51%	4.32%	9.3	6.8	Jan 06
Other Equity	Inst.	14.6	1.8	33.3	367.9	7.26%	5.58%	5.1	6.5	Sep 09
	Retail	42.5	25.3	132.9	240.9	8.27%	7.27%	8.8	6.6	Apr 96
Industrials Sector Equity	Inst.	6.0	5.7	31.7	16.8	7.59%	6.41%	6.6	9.0	Jan 06
	Retail	29.2	23.9	123.6	188.6	8.96%	10.50%	8.9	7.6	Oct 01
Australia & New Zealand Eq.	Inst.	4.9	1.0	73.5	56.7	14.47%	13.70%	7.8	4.0	Dec 08
	Retail	10.0	1.9	149.3	42.1	8.03%	8.92%	8.0	3.8	Feb 05
Canada Fixed Income	Retail	88.9	12.9	296.2	1'064.0	4.91%	5.94%	9.5	6.5	Jan 92
Singapore Equity	Inst.	3.4	1.0	61.6	13.4	15.31%	14.31%	5.5	6.4	Apr 09
	Retail	7.6	3.0	110.9	159.1	10.64%	10.25%	7.4	8.9	May 02
Canadian Eq. Mid/Small Cap	Retail	81.1	2.8	179.6	111.6	5.84%	1.42%	7.5	5.4	Apr 07
Taiwan Equity	Inst.	3.0	1.0	23.6	4.3	37.40%	29.26%	1.7	2.3	Dec 08
	Retail	135.2	9.9	55.9	363.3	10.42%	9.17%	11.1	7.4	Jul 03
Australia Fixed Income	Retail	6.7	1.2	195.8	26.5	3.24%	3.96%	2.8	3.8	Jun 10
Malaysia Fixed Income	Retail	52.9	1.0	64.5	205.8	1.81%	0.75%	6.5	9.3	Oct 07