



DEPARTMENT OF ECONOMICS  
AND BUSINESS ECONOMICS  
AARHUS UNIVERSITY



Center for Research in Econometric Analysis of Time Series

## **Picking Funds with Confidence**

**Niels S. Grønberg, Asger Lunde, Allan Timmermann  
and Russ Wermers**

**CREATES Research Paper 2017-13**

# Picking Funds with Confidence\*

Niels S. Grønborg  
*Aarhus University*

Allan Timmermann  
*UCSD*

Asger Lunde

*Aarhus University*

Russ Wermers

*University of Maryland*

February 13, 2018

## Abstract

We present a new approach to selecting active mutual funds that uses both portfolio holdings and fund return information to eliminate funds with predicted inferior performance through a sequence of pairwise fund comparisons. Our methodology determines both the number of skilled funds and their identities; funds identified as being superior earn substantially higher risk-adjusted returns than top funds identified by conventional alpha ranking methods. Importantly, we find strong evidence of time-series variation in the number of funds identified as superior, as well as fluctuations in the style and industry exposures of such funds over time and across different volatility states.

Key words: Fund confidence set; equity mutual funds; risk-adjusted performance

JEL codes: G2, G11, G17

---

\*We thank conference participants at the 2016 SoFiE conference and seminar participants at Erasmus University, Georgetown University, Purdue University, Texas A&M University, University of Illinois at Urbana-Champaign, University of Maastricht, Tilburg, and CREATES for comments on the paper. Grønborg gratefully acknowledges support from the The Danish Council for Independent Research, Social Sciences (4091-00190/FSE). Grønborg, Lunde, and Timmermann are research fellows at CREATES. Grønborg is a research fellow at the Danish Finance Institute.

# 1. Introduction

Each year, large sums of money and vast amounts of effort are dedicated towards selecting the best mutual funds.<sup>1</sup> Despite such interest, and despite decades of academic research, precisely identifying skilled active mutual funds remains an elusive goal. While many papers have singled out different characteristics of individual funds that are correlated with performance, it is unclear how to effectively identify the *set* of funds that is most likely to outperform in the future.<sup>2</sup> Indeed, the identification of a superior set of funds, rather than an individual fund, involves a hugely more complicated statistical problem, and one that involves a multiple hypothesis test with numerous reasonable alternative hypotheses.<sup>3</sup>

Consider the problem of selecting a menu of actively managed funds for a Defined-Contribution (DC) plan by a fiduciary (perhaps assisted by an investment consultant). This problem is very common, and affects the investments of millions of workers in the United States. Further, thousands of investment alternatives, including mutual funds, are available for consideration by such a fiduciary. Two main issues plague the identification of the set of funds most likely to provide the best future risk-adjusted performance for such a DC plan, net of fees. First, members of this set can change substantially as broad economic conditions evolve, resulting in the set dramatically expanding or shrinking, as well as containing different component funds during different years.<sup>4</sup> Specifically, the nature of a fund's information—and the fund's strategies for acting on such information—can depend on the competitiveness in the fund's peer-group as well as the state of the economy, both of which evolve and can lead to changes in the set of funds that outperform.<sup>5</sup> For example, the ability of a fund to outperform may be short-lived, as evidence of successful strategies attracts competitors (Hoberg, et al., 2017) as well as additional investor flows.

A second major issue is whether an investor can, *ex-ante*, locate the number and identity

---

<sup>1</sup>As of year-end 2016, over \$4.6 trillion was invested in actively managed U.S. equity mutual funds (see [http://www.icifactbook.org/deployedfiles/FactBook/Site%20Properties/pdf/2017/17\\_fb\\_table42.pdf](http://www.icifactbook.org/deployedfiles/FactBook/Site%20Properties/pdf/2017/17_fb_table42.pdf)). A large industry of investment advisors and consultants is engaged in advising retail and institutional clients on how to select funds. For instance, according to the 2015 Cerulli RIA Marketplace 2015 report, there were more than 56,000 registered investment advisors in the U.S. in 2014.

<sup>2</sup>See, e.g., Cremers and Petajisto (2009), Kacperczyk, et al. (2005, 2008), and Wei, et al. (2015). For a survey of the literature on the characteristics that predict active fund performance, see Jones and Wermers (2011).

<sup>3</sup>See Barras, Scaillet, and Wermers (2010) for a discussion of the complexity of a multiple testing problem in the context of the actively managed mutual fund space.

<sup>4</sup>Recent studies suggest that the ability of individual mutual funds to outperform varies over time. For example, Kacperczyk et al. (2014) find that the investment strategies of mutual funds, as well as their ability to outperform, depend on whether the economy is in an expansion or in a recession state. Similarly, Glode, et al. (2011) find evidence of strong (weak) predictability of mutual fund returns following periods of high (low) market returns. Ferson and Schadt (1996), Avramov and Wermers (2006), and Banegas et al. (2013) show that macroeconomic state variables can be used to better predict the future performance of individual equity mutual funds.

<sup>5</sup>Kacperczyk, et al. (2016) provide a model that delivers predictions motivating empirical tests of the changing size of this set. Their model predicts larger outperformance for skilled funds during stressed states in the economy (e.g., high market volatility), as compared to more stable states.

of the set of active funds that will outperform in the future (with a high level of probability). Estimates of fund risk-adjusted returns (alphas) tend to be surrounded by large sampling errors, as outperforming fund managers tend to carry a greater level of idiosyncratic risk (Kacperczyk, et al., 2005). In addition, the estimation error in using past fund data renders any fund-by-fund analysis as resulting in a large number of “false discoveries” of funds with disappointing future performance. The model of Kacperczyk et al. (2016) implies that dispersion in return performance across active fund managers increases during periods with high uncertainty, which both heightens the need for an efficient identification of the best funds and makes it more statistically challenging.<sup>6</sup>

Mutual funds’ fleeting competitive advantage—and the challenge to identifying transient skills—has been discussed in both empirical and theoretical studies. Empirically, Carhart (1997) finds that the performance of top-ranked funds reverts towards the mean after about one year, while Bollen and Busse (2004) find abnormal performance that lasts for one quarter, but disappears at longer horizons. Berk and Green (2004) propose a theoretical model that leads to fund alphas that converge toward zero. Mamaysky, et al. (2008) develop a model in which managers observe private information signals which revert toward being uninformative. Similarly, Glode et al. (2011) present a flow-based model in which diseconomies-of-scale at the fund level remove any abnormal performance over time as investors allocate more money to small funds with high past alphas and allocate less money to large funds with negative past alphas.<sup>7</sup>

This paper introduces an efficient approach to identifying (ex-ante) the set of funds with superior performance, as well as identifying how large the selected set of funds is, what types of investment strategies they adopt, and how this set of funds (along with its risk-adjusted performance) evolves over time. Our approach to identifying the set of “best” (or superior) funds requires not only that we compare each fund’s performance against a single benchmark (or a set of risk factors, as is common practice), but that we also conduct a large set of pairwise comparisons of all funds in existence to eliminate any funds whose performance is dominated by at least one other fund.<sup>8</sup> This comparative process provides an efficient approach to capturing the set of the most skilled fund managers, as it sequentially eliminates funds from the choice set that are likely to reduce the composite performance of the set.

Besides providing valuable information for investors, DC fiduciaries, and consultants, our approach also sheds new light on the nature of competition in the mutual fund industry. Specifically,

---

<sup>6</sup>That is, the model predicts that, during periods with high aggregate volatility, individual fund managers depend more on their private signals, and less on common, aggregate signals. The result is increased cross-sectional heterogeneity in fund-manager beliefs and, as a consequence, in their investment strategies and portfolio returns.

<sup>7</sup>Conversely, the analysis in Garleanu and Pedersen (2016) suggests that, due to differential search costs, funds populated by the most sophisticated investors can be expected to outperform, in equilibrium.

<sup>8</sup>As we will show, this pairwise comparison is efficient, in part, because it discards a fund when it is judged to be significantly worse than any other fund, and no longer considers it in further evaluations of the remaining set. With traditional multiple hypothesis tests, the full (estimated) covariance matrix must be employed, which requires the estimation of covariance of each fund with every other fund—a monumental task.

by singling out the set of funds with superior performance and showing how this set evolves through time, we are able to analyze how many funds are in this set, for how long they can maintain their competitive edge, and, thus, how rapid turnover is among funds included in the set. Moreover, by cross-checking fund performance against quarterly holdings, we gain information into which strategies the funds pursue in order to obtain superior performance.<sup>9</sup>

Conventional approaches in the finance literature are not well-designed to handle such comparisons, nor do they control for the “size” of the test, i.e., the probability of wrongly eliminating truly superior funds.<sup>10</sup> To deal with such issues, our analysis adopts a new approach for the selection of mutual funds that makes use of the Model Confidence Set (MCS) methodology of Hansen, Lunde, and Nason (HLN; 2011), which, in turn, is designed to select the most accurate prediction models from a large set of candidate models. In our context, the set of “candidate models” is the set of alpha forecasts, one for each equity mutual fund in existence at a given point in time; these individual fund alpha forecasts are generated by a particular risk model (e.g., the four-factor model of Carhart, 1997). Using the MCS methodology, our approach undertakes a series of pair-wise tests to sequentially eliminate funds with inferior performance (i.e., those with an “inferior forecasting model”). Specifically, if at least one fund with significantly inferior performance, relative to *any* other fund, is identified, this fund is eliminated, and the elimination process continues on the reduced set of funds.<sup>11</sup> The procedure continues until no further funds with inferior performance, relative to any other fund, can be identified and eliminated. We label the set of funds remaining at the end—the funds forecasted to have superior performance—as the “Fund Confidence Set” (FCS).

Since the FCS is designed to choose the best among a set of funds using a particular forecasting model, it is crucial that the underlying performance model is sufficiently informative about fund manager skills. To achieve this, as a further contribution of our paper, we propose a novel (yet, parsimonious in its data requirements) performance model that (i) allows for time-varying alphas using the latent skill approach of Mamaysky et al. (2008); and (ii) extracts alpha estimates, pooling information from past fund returns and fund holdings. Empirically, we find that the FCS

---

<sup>9</sup>We note that our search for funds with superior performance uses only simple and widely available data on past returns and publicly reported holdings—both of which are readily available to the public at the time a decision on whether to invest in the fund is made by our empirical tests.

<sup>10</sup>While the “False Discovery Rate” procedure used by Barras, Scaillet, and Wermers (2010) provides an efficient approach to estimating the Type I error (size) associated with identifying a set of outperforming funds, it says little about the error rate in identifying individual funds within a given set that are expected to outperform. Indeed, Barras et al. (2010) apply a simple approach that examines the set of funds with the best estimated alphas (or t-statistics), with no mechanism to winnow out individual funds within this set that are unlikely to be skilled according to a desired level of Type I error. That is, the simplicity of the FDR approach compromises the selection of the most-skilled funds by stopping the analysis at the level of the initially chosen set of funds, e.g., all those having an alpha p-value lower than 5%.

<sup>11</sup>Note that the efficiency of this approach exploits the fact that we wish to generate a set of funds forecasted to have the best performance, and not, e.g., an optimized portfolio of funds with the best mean-variance properties. Thus, our approach can be used as the first step in assembling a portfolio with desired properties (which will vary among different types of investors).

approach can be used to select a set of funds with considerably better average performance than techniques that employ a fixed proportion (e.g., 5% or 10%) of top ranked funds. In particular, an equal-weighted portfolio of funds included in the top FCS generates a four-factor alpha of almost 50 bps/month (which is highly statistically significant); by comparison, the four-factor alpha of a constant allocation to the top decile of funds (i.e., as employed by Carhart, 1997) is only about 2 bps/month and is statistically insignificant.

Because the performance of most funds is estimated with large sampling errors, in part because most funds have a limited history, our approach is found to work best if we use a relatively stringent criterion for inclusion of funds in the FCS—resulting in a relatively narrow set of funds with superior performance and a reduced probability of wrongly including funds with inferior performance. Using this stringent approach, we find that the set of funds identified as superior fluctuates considerably over time. Sometimes, only a few funds (or even a single fund) are selected; at other times, the approach identifies a large set of funds as being superior. As we relax this stringent inclusion criterion, we still find that the FCS identifies a superior set of funds, but with slightly weaker economic and statistical significance. In further tests, we find that the size of the set of funds with (expected) superior performance is found to be significantly correlated with a range of macroeconomic state variables, as the fraction of superior funds is larger in expansions, and smaller in recessions, suggesting that funds’ ability to outperform is state-dependent.<sup>12</sup>

We also find that dispersion in fund performance is greater during times with high levels of market volatility, leading to the best funds—as identified by our FCS approach—performing better during such periods, relative to calmer times. This is consistent with the model of Kacperczyk, et al. (2016), which suggests that the rewards to skilled investors from processing information are higher when market risk is high. By comparison, we find no significant following-year outperformance of top (simple alpha-ranked) decile funds during high risk periods. This result highlights the efficacy of our risk model, which carefully eliminates funds that mainly contribute idiosyncratic risk—and not alpha—to the set of superior funds.

In a deeper analysis of the portfolio holdings data of the mutual funds, we find that the evolving set of funds that are identified as superior results in an implied time-series change in industry concentration and risk loadings. For example, superior FCS funds tilt toward growth stocks from 1994 to 2003 (a period of great return variability in the cross-section of growth stocks), as well as from 2009 to 2011. Conversely, they overweight value stocks from 2003 to 2009 (a period of great return variability in the cross-section of value stocks). Superior funds also overweight small stocks from 2001 to 2003 and 2009 to 2012, but overweight large capitalization stocks from 2005 to 2008. Superior funds have above-average exposures to the momentum factor only during

---

<sup>12</sup>This result highlights the difficulty in identifying, ex-ante, superior funds in a recession setting—even though the average performance of active funds is better during recessions—due to the short time-series nature of recessions. For instance, there are only 34 months of NBER-identified recessions during our entire sample period.

brief spells. Turning to industry concentration, funds in the top FCS overweight computer and electronic equipment stocks after 2005, and also overweight business services and machinery stocks during shorter spells. Conversely, they underweight retail stocks and, in particular, banking stocks throughout most of the sample. All of these results indicate that the identification of superior funds is both time-varying and complex; our modeling procedure uniquely provides an approach to such identification.

The FCS methodology can also be used to successfully identify funds with inferior performance. When applied to select inferior funds, we find that the set of “worst” funds is somewhat wider than the equivalent set of superior funds. This reflects the greater persistence of factors giving rise to underperformance, such as high trading costs and management fees (relative to skills). Again, we find that the funds in the bottom FCS portfolio produce substantially worse performance than a portfolio consisting of a fixed proportion of alpha-ranked funds, such as the bottom 5-10% of funds, indicating that the FCS effectively identifies funds with poor skills rather than funds that have simply experienced a spell of bad luck.

Our analysis differs from previous studies in several important dimensions. Kosowski, et al. (2006) ask whether there exist star fund managers, i.e., if the single manager with the best estimated alpha has truly outperformed her benchmark model, such as the four-factor model of Carhart (1997), or, alternatively, if a single fund at a particular ranked fractile (i.e., the fund at the 10 percentile point), can outperform. However, their methodology cannot be used to endogenously determine the size of the *set* of funds with superior performance or the identity of the individual funds, a much tougher econometric problem. In fact, a strategy of only investing in the fund whose alpha is deemed highest can sometimes backfire, because such a fund might have been “lucky,” and its performance could reflect very high idiosyncratic risk taking. Barras, et al. (2010) develop an approach that controls for funds whose high alpha estimates are due to “luck.” However, unlike us, they do not address the identification of *individual* funds deemed to have the highest (positive) alphas. This is a highly relevant question from an investor’s perspective, because not all funds, even those with truly positive alphas, should be expected to perform equally well.<sup>13</sup> In addition, and of great relevance to a fiduciary charged with selecting a set of active funds, individual investors typically wish to invest in only one or a few funds, and not a larger set. Thus, the identification of the set of superior funds is of great value to such investors. Further, Barras, et al. (2010) find evidence that the set of funds with positive alphas has been shrinking over time, making it more difficult to identify truly superior fund managers. This result highlights the importance of identifying individual funds that are predicted to outperform in order to form a precise set, rather than starting with a set and conducting inference on the entire set (equal-weighted) as is done by

---

<sup>13</sup>In fact, we show that an approach similar to that of Barras et al. (2010) can be used to screen the initial set of candidate funds from which the fund confidence set is selected.

Barras, et al. (2010).<sup>14</sup>

We show empirically that our approach, which endogenously builds the set of funds (starting with pairwise individual fund comparisons) expected to produce superior performance, is far better at identifying truly skilled managers than the conventional portfolio sorting approach. Thus, our approach has two major advantages over past methods. First, it helps to discard individual funds that lie in the top past-performance fractile due to luck. And, second, our approach helps to adjust the search for superior active managers in accordance with the breadth of active management skills during a particular period of time, broadening or narrowing the set of chosen funds as market conditions evolve. We show that these two advantages translate into superior fund alphas, as well as better identification of inferior funds. The key to our approach is its efficient pairwise comparisons of funds to separate luck from skill.

The outline of our paper is as follows. Section 2 introduces the fund confidence set methodology, while Section 3 describes our data and the model used to measure the performance of individual funds. Section 4 presents performance results for the FCS portfolios and Section 5 provides details on which funds get selected to be among the superior or inferior funds and how this set varies through time. Section 6 performs an attribution analysis which decomposes the performance of the FCS funds using holdings data on industry concentrations and stock-level estimates of individual funds' exposures to common risk factors. Section 7 concludes.

## 2. The Fund Confidence Set

A large empirical literature in finance explores whether it is possible, *ex ante*, to identify funds with superior risk-adjusted performance. The most widespread practice used in the literature is, first, to rank individual funds based on their forecasted performance; second, to form decile portfolios based on such rankings; and, third, to track the subsequent risk-adjusted performance of the portfolios of funds. While the practice of allocating individual funds into decile portfolios is simple to perform and intuitive to interpret, it only addresses whether the *average* risk-adjusted performance of the top 10% of funds (or a similar proportion) is positive. This question is very different from the more relevant and interesting question of whether we can identify a particular set of individual funds with superior performance, regardless of whether they reside in a particular past-performance ranking quantile.<sup>15</sup> For example, if the set of funds capable of producing positive risk-adjusted performance

---

<sup>14</sup>Harvey and Liu (2016) estimate individual fund alphas using a random effects panel approach that pools time series and cross-sectional information from the population of funds. This approach does not endogenously identify the set of funds with superior performance, however.

<sup>15</sup>An especially salient example of the failure of simple ex-post ranking approaches is the “stars” system of Morningstar. Kinnel (2010) shows that expense ratios, by themselves, are better predictors of future fund performance than the Morningstar star rating—which includes the effect of both pre-expense past returns and current expense ratios (both measured relative to peers). Morey (2003) finds that Morningstar five-star rated funds experience a severe drop in performance over the following three years.



varies over time, and is sometimes far narrower than 10%, then we may well find, empirically, that the top 10% of funds do not generate positive performance, on average, even though there exists a set of funds with positive (ex-ante) risk-adjusted performance. Or, suppose that the top 10% of funds produce positive risk-adjusted returns, but that the top 5% produce much higher (and more precisely estimated) risk-adjusted returns. In such a case, an investor would be well-served to invest only in the top 5% of funds in order to maximize her chances of generating an economically significant level of outperformance. As a more complicated example, suppose that the best funds are dispersed through the top 10% alpha “tail”. In this case, the investor would like to know the identities of the most skilled funds, and not simply the identity of all funds in the tail.<sup>16</sup>

This section introduces an alternative approach to fund selection that does not fix the proportion of funds deemed capable of delivering superior performance but, rather, determines this proportion endogenously, as part of the process used to estimate the performance of individual funds. Our econometric approach reflects how a sophisticated investor with access to historical data on individual fund returns and periodic security holdings would go about identifying funds with skills and thus mimics the search process underlying the analysis in papers such as Garleanu and Pedersen (2016). We first describe the approach in broad terms and characterize its properties, then provide details on how we implement the approach on our mutual fund data.

Before formally introducing our Fund Confidence Set approach, we provide a motivating example in which the set of funds is small enough that we can gain intuition for how the approach works.

### *2.1. Selecting Funds with Superior Performance: A Motivating Example*

Understanding the properties of different approaches for fund selection can be quite complex in situations with hundreds of funds in existence. To keep the discussion simple and intuitive and to illustrate the differences across different approaches for fund selection, we first consider an example with only four funds, namely Endowments Inc., Seligman Communications, Fidelity Selected Technology, and Fidelity Selected Electronics. We assume that an investor has obtained alpha estimates for these funds based on data available in July 2002 and wants to determine which, if any, of these funds to hold.

A simple way to select funds would be to look at their average alphas. For the four funds under consideration, alpha estimates are presented in Panel A of Table 1. Because all estimates are positive, the four funds could be regarded as worthy of investment as indicated by the checkmarks in the table. This logic ignores sampling errors, however, and fails to assess the statistical significance of the alpha estimates. To account for sampling error, the investor could instead test

---

<sup>16</sup>As motivation, Barras, et al. (2010) find that the estimated percentage of skilled funds in the right tail varies substantially over time, and that skilled managers are sometime more dispersed (mixed with unskilled but lucky managers) in the right tail than at other times.

the statistical significance of the alpha estimates by means of a simple  $t$ -test. The values of the associated  $t$ -statistic are presented for each fund in Panel B of Table 1. Using conventional levels of significance, Endowments Inc., Fidelity Selected Technology, and Fidelity Selected Electronics would be identified as having abnormally good performance.

The advantage of using the  $t$ -statistics for the alpha estimates rather than the alpha estimates themselves is that  $t$ -statistics account for the uncertainty in estimation of the alphas.  $T$ -statistics are not well suited, however, for comparing the performance across different funds, as they ignore the correlation in performance between different funds. In other words, they do not address the possibility that two funds have significantly positive alpha estimates, but one fund clearly dominates the other. In this situation, it may not be optimal to invest in both funds.

To address this possibility and distinguish between superior, neutral and inferior funds, one can conduct pairwise comparisons of the funds' performance. If, for example, two funds have a very similar focus, the difference in their average alpha might be very small and this could lead to similar  $t$ -statistics. However, if two funds produce different alphas and have very similar risk exposures we would not want to label both as superior, even if both alphas are significantly positive. In fact, a high correlation between two funds' returns (induced by similar risk exposures) should help us in identifying situations where one fund dominates the other. Panel C of Table 1 presents the outcome of a stepwise procedure for comparing the performance of the four funds under consideration.<sup>17</sup> In Step 1, all six pairwise  $t$ -statistics for the four funds are presented. The largest test statistic (reported, in the table, for the funds in the rows relative to the funds in the columns) stems from comparing Fidelity Selected Technology and Endowments Inc.<sup>18</sup> Using a bootstrap methodology described below, we conclude that the performance of Endowments Inc. is significantly worse than that of Fidelity Selected Technology, and we, therefore, eliminate Endowments Inc.<sup>19</sup> In Step 2, we consider the three test statistics for pairwise comparisons of the remaining funds. The largest test statistic arises from the comparison of Fidelity Selected Electronics and Seligman Communications. Using bootstrapped critical values, we can eliminate Seligman Communications. In Step 3 we only have one comparison left. The comparison of the two Fidelity funds results in a test statistic of -0.66, which would not lead to elimination of any of the remaining funds at most conventional levels of significance.

Note from this example that Seligman Communications, which has a positive but statistically insignificant alpha, could not be eliminated based on a pair-wise comparison with Fidelity Selected Technology alone. Instead this fund was eliminated through its pairwise comparison with Fidelity Selected Electronics Likewise, Endowments Inc., which has a significantly positive alpha, could only be eliminated after a pairwise comparison with Fidelity Selected Technology. The fact that

---

<sup>17</sup>This stepwise approach is similar to what we will be using in our subsequent analysis.

<sup>18</sup>The tests use the range statistic of Hansen et al. (2011), which we discuss further below.

<sup>19</sup>A key contribution of Hansen et al. (2011) is to calculate  $p$ -values for all the pairwise test statistics that remain valid in spite of the stepwise comparisons.

Endowments Inc. can be eliminated based on a comparison with Fidelity Selected Technology (but not based on a comparison with Fidelity Selected Electronics) is explained by the higher average alpha of Fidelity Selected Technology as the two Fidelity funds have very similar correlations with Endowment Inc. Conversely, the fact that Seligman Communications can be eliminated based on a comparison with Fidelity Selected Electronics and not through a comparison with Fidelity Selected Technology is explained by the difference in their return correlations. Fidelity Selected Electronics is much higher correlated with Seligman Communications than Fidelity Selected Technology is. Hence, this example shows that both the absolute (alpha) performance, but also the correlation in returns matter to our ability to identify funds with superior performance.

In particular, the process of pairwise comparisons can be extremely demanding for large sets of funds. The simple example presented here suggests that some funds could be removed in an initial step based on the test statistics in Panel B of Table 1. In our example, this would entail eliminating Seligman Communications based on its low test statistic from Panel B and skipping Step 1 in Panel C. In our analysis we rely on such an initial step, where funds with insignificant alphas are eliminated based on the methodology of Romano and Wolf (2005). This is because it is very unlikely that funds with the lowest individual t-statistics will be superior to other funds, when compared on a pairwise basis. Such an outcome would require that such a fund have a much lower correlation with every other fund, as compared to the pairwise correlations of all other funds with each other—a very unlikely event.

## 2.2. Predictive Alpha

For a fund to be attractive to investors, it must have a high expected risk-adjusted performance, relative to the benchmark model that those investors use. This requires that the fund’s performance be at least modestly predictable.<sup>20</sup> Our objective is, therefore, to identify funds for which, with some confidence, we forecast positive future risk-adjusted returns. Following common practice, we compute a fund’s risk-adjusted return by adjusting the fund’s excess returns, net of the T-bill rate,  $R_{i,t}$ , for its exposure to a set of risk factors,  $\mathbf{z}_t$ :

$$R_{i,t} = \alpha_{i,t} + \beta'_{i,t}\mathbf{z}_t + \varepsilon_{i,t}. \quad (1)$$

Here  $i$  refers to the fund and  $t$  refers to the time period;  $\varepsilon_{i,t} \sim (0, \sigma_{\varepsilon_i}^2)$  is the fund’s idiosyncratic return,  $\beta_{i,t}$  measures the fund’s exposure to the common risk factors, while  $\alpha_{i,t}$  measures its risk-adjusted (abnormal) performance. The model in equation (1) is quite general as it allows both  $\alpha_{it}$

---

<sup>20</sup>To see the importance of this point, consider a fund that takes high idiosyncratic risks and has produced a high average (historical) risk-adjusted performance due to a single high monthly return outlier. This would not instill much confidence in the fund’s ability to produce high future returns. Contrast this with another fund with a lower average risk-adjusted return but consistently good performance; this fund is likely to be more attractive to investors, particularly if the periods when it outperforms can be predicted.

and  $\beta_{it}$  to vary over time. If a fund's alpha is constant over time, the fund's average historical performance can be used to compute its expected future performance. Conversely, if a fund's abnormal performance changes over time, we need to model how such changes evolve.

Given an estimate of the expected value of fund  $i$ 's alpha during period  $t$  based on information available at time  $t - 1$ ,  $\hat{\alpha}_{i,t|t-1}$ , we need to determine if  $\hat{\alpha}_{i,t|t-1}$  can reliably predict whether the fund will outperform in the future. To this end, we consider the following "predictive alpha":

$$P_{i,t} = \text{Max}(\hat{\alpha}_{i,t|t-1}, 0) \times \text{sign}(R_{i,t} - \hat{\beta}'_{i,t}\mathbf{z}_t). \quad (2)$$

Here the sign function  $\text{sign}(\bullet)$  equals +1 if the argument is positive, -1 if the argument is negative, or 0 if the argument is zero. To motivate the objective in (2), note that  $P_{i,t}$  will be large for funds with large, positive predicted alphas ( $\hat{\alpha}_{i,t|t-1} > 0$ ) whose subsequent risk-adjusted returns are positive ( $R_{i,t} - \beta'_{i,t}\mathbf{z}_t > 0$ ). Conversely, the objective in (2) penalizes funds with large positive alpha estimates but negative future risk-adjusted returns. Taking the expectation of  $P_{i,t}$ , it is easy to see that the expected predictive alpha in (2) is the forecasted alpha, conditional on the forecast being positive, weighted by the difference in probability that the conditional forecast is directionally correct vs. incorrect.<sup>21</sup> The predictive alpha also preserves scale, i.e., it is measured in units of returns (time<sup>-1</sup>). The above predictive alpha excludes funds with negative predicted alphas ( $\hat{\alpha}_{i,t|t-1} < 0$ ), since we wish to locate only positive alpha funds in our first application. We will identify funds with inferior forecasted performance using an analogous procedure in Section 4.5 below.

To select funds with the highest expected value of the predictive alpha using available past sample information, for each fund,  $i$ , we measure the past success in predicting this fund's performance up to time  $t$ , using the sample estimate of the average value of (2):

$$\bar{P}_{i,t} = \frac{1}{t - t_{i0}} \sum_{\tau=t_{i0}+1}^t P_{i,\tau} = \frac{1}{t - t_{i0}} \sum_{\tau=t_{i0}+1}^t \text{Max}(\hat{\alpha}_{i,\tau|\tau-1}, 0) \text{sign}(R_{i,\tau} - \hat{\beta}'_{i,\tau}\mathbf{z}_\tau). \quad (3)$$

Here,  $\hat{\alpha}_{i,\tau|\tau-1}$  is the forecast of  $\alpha_{i,\tau}$  based on information available at time  $\tau - 1$ , and  $\hat{\beta}'_{i,\tau}$  are least-squares estimates of  $\beta'_i$ , using data only up to time  $\tau$ . Finally,  $t_{i0} + 1$  is the starting point of the sample used to estimate the  $i$ th fund's predictive alpha performance up to time  $t$ ,  $\bar{P}_{i,t}$ .

---

<sup>21</sup>Note that this function imposes a larger reward (penalty) on forecasts that are "more correct" ("more incorrect"), i.e., that predict both the right (wrong) sign and predict it with greater confidence (through the magnitude of the term  $\text{Max}(\hat{\alpha}_{i,t|t-1}, 0)$ ).

### 2.3. Fund Confidence Set

Our data contain more than 2,000 funds whose performance needs to be pair-wise compared at each point in time. This introduces a complicated multiple hypothesis testing problem, which we address by applying the model confidence set (MCS) approach of HLN. That approach is designed to choose the set of “best” forecasting models from a larger set of candidate models. Because the approach is developed for the selection of forecasting models, we need to modify it for our setting. Most obviously, the object of interest in our analysis is not a model, but a model applied to a fund’s return performance, and, so, we label our approach the Fund Confidence Set (FCS). We next describe how the approach works.

Our goal is to select a set of funds which, at a certain level of confidence, contains the best fund—or set of funds, if multiple funds are believed to have identical performance. The approach relies on an equivalence test and an elimination rule. Let  $\mathcal{F}_t^0 = \{F_{1t}, \dots, F_{nt}\}$  be the initial set of funds under consideration while  $P_{it}$ , given by equation (2), measures fund  $i$ ’s performance in period  $t$ . The difference between the performance of funds  $i$  and  $j$  at time  $t$  is then

$$d_{ij,t} = P_{i,t} - P_{j,t}, \quad i, j \in \mathcal{F}_t^0. \quad (4)$$

Defining  $\mu_{ij} = E[d_{ij,t}]$  as the expected difference in the performance of funds  $i$  and  $j$ , we prefer fund  $i$  to fund  $j$  if  $\mu_{ij} > 0$ ; both funds are judged to be equally good if  $\mu_{ij} = 0$ . The set of “superior” funds at time  $t$ ,  $\mathcal{F}_t^*$ , consists of those funds that are not dominated by any other funds in  $\mathcal{F}_t^0$ , i.e.,  $\mathcal{F}_t^* = \{i \in \mathcal{F}_t^0 : \mu_{ij} \geq 0 \text{ for all } j \in \mathcal{F}_t^0\}$ . The FCS approach identifies  $\mathcal{F}_t^*$  by means of a sequence of tests, each of which eliminates the worst fund—the fund deemed to be worst relative to another fund in the current set of surviving funds,  $\mathcal{F}_t$ —provided that this difference is statistically significant. Each round of this procedure tests the null hypothesis of equal performance

$$H_{0,\mathcal{F}_t} : \mu_{ij} = 0, \text{ for all } i, j \in \mathcal{F}_t \subset \mathcal{F}_t^0, \quad (5)$$

against the alternative hypothesis that the expected performance differs for at least two funds:

$$H_{A,\mathcal{F}_t} : \mu_{ij} \neq 0 \text{ for some } i, j \in \mathcal{F}_t. \quad (6)$$

Following HLN, we define the Fund Confidence Set (FCS) as any subset of  $\mathcal{F}_t^0$  that contains  $\mathcal{F}_t^*$  with a certain probability,  $1 - \lambda$ , where  $\lambda$  is the size of the test.

With these definitions in place, we next explain how the algorithm for constructing the FCS works. The first step sets  $\mathcal{F}_t = \mathcal{F}_t^0$ , the full list of funds under consideration at time  $t$ . The second step uses an equivalence test to test  $H_{0,\mathcal{F}_t} : E[d_{ij}] = 0$  for all  $i, j \in \mathcal{F}_t^0$  at a critical level  $\lambda$ . If  $H_{0,\mathcal{F}_t}$

is accepted, the FCS is  $\hat{\mathcal{F}}_{1-\lambda,t}^* = \mathcal{F}_t$ . If, instead,  $H_{0,\mathcal{F}_t}$  gets rejected, the elimination rule ejects one fund from  $\mathcal{F}_t$ , and the procedure is repeated on the reduced set of funds. The procedure continues until the equivalence test does not reject, and, so, no additional funds need to be eliminated. The remaining set of funds in this final step is  $\hat{\mathcal{F}}_{1-\lambda,t}^*$ .

Elimination of individual funds is based on their relative sample performance, using  $\bar{P}_{i,t}$  from (3). Specifically, from (3) and (4), we estimate the performance of fund  $i$  relative to fund  $j$  as  $\bar{d}_{ij} = t^{-1} \sum_{\tau=1}^t \hat{d}_{ij,\tau}$ . To obtain a better behaved test statistic, we divide this measure by its standard error,  $\sqrt{\widehat{\text{var}}(\bar{d}_{ij})}$ , to obtain<sup>22</sup>

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{\text{var}}(\bar{d}_{ij})}}, \quad (7)$$

As in HLN, we base the test of  $H_{0,\mathcal{F}_t}$  on the range statistic

$$T_{\mathcal{F}_t} = \max_{i,j \in \mathcal{F}_t} |t_{ij}|, \quad (8)$$

which finds the largest  $t$ -statistic chosen among the many pairwise  $t$ -tests in (7).

Under assumptions listed in HLN, the set of pairwise  $t$ -tests,  $t_{ij}$ , are asymptotically jointly normally distributed with unknown covariance matrix,  $\Omega$ . Because so many pairwise test statistics are being compared and  $\Omega$  is unknown, the resulting test statistic,  $T_{\mathcal{F}_t}$ , has a non-standard asymptotic distribution whose critical values can be bootstrapped using the approach of White (2000). From these draws, the following sequential elimination rule is used to identify the fund ( $i$ ) with the worst estimated performance measured relative to some other fund ( $j$ ) as the candidate for elimination.<sup>23</sup>

$$\arg \min_{i \in \mathcal{F}_t} \sup_{j \in \mathcal{F}_t} t_{ij}. \quad (9)$$

Based on (8) and (9), any fund whose performance looks sufficiently poor, relative to that of at least one other fund currently included in the FCS is purged.<sup>24</sup>

Because a random number of possibly dependent tests are carried out, it is far from trivial to control the coverage probability of this stepwise procedure. Notably, if each round conducts a test at a fixed critical level,  $\lambda$ , then the final FCS will have a very different coverage probability than  $1 - \lambda$ . A key contribution of HLN is to design a sequential procedure that can be used to control the

<sup>22</sup>Also, by scaling  $\bar{d}_{ij}$  by its standard error, the resulting test statistic is pivotal and can be expected to have better sampling properties.

<sup>23</sup>Our objective is to eliminate funds for whom our risk model has poor forecasting power, while HLN's objective is to eliminate models with the greatest level of mispricing. Accordingly, we alter the elimination rule, as compared to HLN, who use  $\arg \max_{i \in \mathcal{F}_t} \sup_{j \in \mathcal{F}_t} t_{ij}$ .

<sup>24</sup>Specifically, if the FCS  $p$ -value for the fund identified by (8) is smaller than the  $\lambda$ -quantile of the bootstrapped distribution, then this fund is deemed inferior to at least one other fund and is eliminated.

coverage probability,  $1 - \lambda$ , of the FCS by dynamically adjusting the bootstrapped critical values employed in each elimination step.<sup>25</sup>

#### 2.4. Choosing $\lambda$

As with any inference problem, the FCS approach requires us to trade off type I and type II errors. Type I errors (false positives) are incorrect rejections of a true null, i.e., wrongly eliminating funds whose performance is as good as that of the best fund. Type II errors, conversely, are failures to reject a false null hypothesis, i.e., failing to exclude a poor fund from the FCS. How these errors are traded off gets regulated by the choice of the size of the equivalence test ( $\lambda$ ) used by the FCS approach, which, therefore, is an important parameter.

Setting  $\lambda$  high means reducing the probability of wrongly including inferior funds (i.e., increasing the power of the equivalence test), but also implies that we stand a reduced chance of including funds with truly superior performance. Conversely, setting  $\lambda$  low means increasing the probability of including both truly inferior and truly superior funds as we become more cautious about eliminating individual funds, and the algorithm becomes less selective.

If the estimated performance of many of the funds is quite noisy, tests of the equivalence hypothesis in (5) may not be very powerful, and the algorithm will eliminate too few funds, resulting in a bloated set that includes many inferior funds.<sup>26</sup> This would simply reflect that the data are not sufficiently informative to distinguish between the performance of different funds. We can easily imagine economic states in which this would plausibly be the case, such as the economic crisis of 2008/2009, where market volatility and fund idiosyncratic volatility were quite high. Conversely, when the data are informative and allow for sharper inference, the equivalence tests first eliminate the poor funds before questioning the superior funds.

In trading off these effects, our goal is to create a set of funds with the highest level of future composite risk-adjusted performance, with a constraint of reducing, to the largest extent possible, the risk of underperformance. We believe that this goal is especially useful for retirement plan fiduciaries, who seek to provide, with a high level of statistical confidence, a menu of the most promising funds for their investors, as well as fiduciaries who select managers for pension funds, endowments, sovereign wealth funds, and other trusts.<sup>27</sup>

We, therefore, opt for a relatively high value of  $\lambda$ , choosing  $\lambda = 0.90$  as our benchmark. However, to illustrate how sensitive our results are to our particular choice of  $\lambda$ , in robustness tests, we also

---

<sup>25</sup>Specifically, Theorem 1 in Hansen et al. (2011) establishes conditions under which the probability that a truly superior fund is captured in the estimated FCS is greater than or equal to  $1 - \lambda$ , while the probability of wrongly including an inferior fund in  $\hat{\mathcal{F}}_{1-\lambda,t}^*$  asymptotically goes to zero.

<sup>26</sup>In other words, funds can avoid being eliminated from the FCS either because they have a high average performance which is precisely estimated, or, alternatively, because their performance is imprecisely estimated so their alpha is surrounded by large standard errors which reduces  $t_{ij}$  in (7).

<sup>27</sup>Such fiduciaries generally wish to minimize Type 2 error (the probability of including an unskilled fund) at the expense of allowing higher Type 1 error (the probability of not including a skilled fund).

consider three alternative values ( $\lambda = 0.75, 0.50, 0.10$ ) which result in fewer funds being eliminated. We refer to the four sets of  $\lambda$ -values as tight ( $\lambda = 0.90$ ), medium-tight ( $\lambda = 0.75$ ), medium ( $\lambda = 0.50$ ), and wide ( $\lambda = 0.10$ ).

### 2.5. Choice of Candidate Set of Funds

We implement the FCS approach as follows.<sup>28</sup> First, because there are more than 2,000 funds in our sample, it is not feasible to conduct all possible pairwise performance comparisons.<sup>29</sup> To handle this issue, we propose a pre-screening step that first (before implementing the FCS) excludes funds with poor or unpredictable performance—either is detrimental from an investment perspective. This step greatly reduces the complexity of the subsequent FCS computation step. Specifically, we consider a strategy that endogenously chooses the initial set of funds ( $\mathcal{F}_t^0$ ) by identifying funds whose predictive alpha estimates are significantly higher than zero. This is accomplished using the stepwise procedure of Romano and Wolf (2005) which aims to identify as many funds with genuinely positive alphas as possible, while controlling the familywise error (FWE) rate. The null and alternative hypotheses tested by the Romano-Wolf approach are  $H_{0i} : \alpha_i \leq 0$  vs.  $H_{1i} : \alpha_i > 0$ , and the FWE rate is defined as the probability of wrongly rejecting the null for at least one fund, i.e.,  $FWE = \text{prob}(\text{rejecting at least one } H_{0i} \text{ for which } \alpha_i > 0)$ . The challenge is to design an approach whose asymptotic FWE is no greater than some critical level, while accounting for the multiple hypothesis testing problem arising from comparing so many mutual funds.<sup>30</sup>

## 3. Performance Measurement Model

This section introduces our approach to estimate the risk-adjusted performance of individual mutual funds. In practice, the performance measurement model plays an important role in the empirical analysis. The FCS approach—or any other approach that relies on conditioning information about fund performance—cannot be expected to work well without informative signals about individual

---

<sup>28</sup>We estimate the fund confidence set using the MulCom 3.0 package for Ox; see Hansen and Lunde (2010) and Doornik (2006).

<sup>29</sup>For example, a cross-section of 2,200 funds observed over 270 months gives rise to just over 650 million pairwise comparisons.

<sup>30</sup>To implement the Romano-Wolf procedure, define a test statistic,  $test_i$ , for testing  $H_{0i}$  vs.  $H_{1i}$ . Suppose we have renumbered the funds  $i = 1, \dots, n$  by the magnitude of their individual test statistics,  $test_n \leq test_{n-1} \leq \dots \leq test_1$ . A critical value,  $c_1$ , is then determined such that the set of  $n_1$  funds with test statistics  $test_{R_1} \geq c_1$  has a fixed coverage probability. Only funds with higher test statistics (i.e., funds numbered  $1, \dots, n_1$ ) are set aside in this step. Next, the procedure is repeated on the remaining  $n - n_1$  funds, resulting in a new critical value,  $c_2$ , and we set aside those funds whose test statistics exceed  $c_2$ . The procedure is repeated until we can no longer reject the null of no superior performance for any remaining funds, at which point the algorithm stops and consolidates the funds set aside in previous steps. We use the studentized StepM method (Algorithm 4) of Romano and Wolf (2005) in the analysis. Like the MCS approach, the Romano-Wolf method uses the White (2000) bootstrap to calculate the critical values used for eliminating funds.



funds' future performance. As we show below, a conditional alpha approach that uses both returns and holdings data can provide sharp inference about alphas than an approach which uses only returns data. Sharper inference on alphas translates into an improved ability to discriminate between truly skilled and truly unskilled fund managers.

In common with much of the existing literature on mutual fund performance, we use a four-factor model that, in addition to the market factor, adjusts for the size and value factors of Fama and French (1992) and the momentum factor of Carhart (1997). However, we generalize this model in two ways. First, following Mamaysky et al. (2007, 2008), we assume that managers receive information (unobserved to the econometrician) that is correlated with future returns. As we show below, such information gives rise to a time-varying component in fund performance. Second, we generalize this framework to combine information from past returns with holdings data to more accurately extract an estimate of fund performance.

The benchmark four-factor model takes the form

$$R_{it} = \alpha_i + \beta'_i \mathbf{z}_t + \varepsilon_{it}. \quad (10)$$

Following common practice in the finance literature, we obtain estimates of  $\theta_i = (\alpha_i \beta'_i)'$  using a rolling 60-month estimation window. Such estimates account for slowly-evolving shifts in mutual fund performance and risk exposures.<sup>31</sup>

### 3.1. Time-varying Skills

Different approaches have been suggested for capturing time variation in fund alphas. We follow the approach of Mamaysky et al. (2007, 2008) and show how it can be generalized—and improved—to take advantage of information from mutual fund holdings data.

Our starting point is a four-factor model. Recall that  $R_{it}$  is the monthly return on fund  $i$ , measured in excess of a 1-month T-bill rate. Similarly, let  $\mathbf{z}_t = (R_{mt}, SMB_t, HML_t, MOM_t)'$  denote the values of the four risk factors, where  $R_{mt}$  is the return on the market portfolio in excess of a 1-month T-bill rate,  $SMB_t$  and  $HML_t$  are the small-minus-big equity market capitalization and value-minus-growth factors of Fama and French (1992), and  $MOM_t$  is the momentum factor of Carhart (1997), constructed as the return differential on portfolios comprising winner versus loser stocks, tracked over the previous 12 months.

Individual funds' alphas are value-weighted averages of their stock-level alphas, so our analysis starts with the performance of individual stocks. Specifically, we decompose the excess return of each stock,  $r_{jt}$ , into a risk-adjusted return component,  $\alpha_{jt}$ , a systematic return component obtained

---

<sup>31</sup>Empirically, we find that five-year rolling window estimates are slightly better at identifying and predicting fund performance than estimates based on an expanding estimation window. The results reported below are, however, robust to using an expanding estimation window.

as the product between a set of risk exposures,  $\beta_j$  and factor returns,  $\mathbf{z}_t$ , and an idiosyncratic return component,  $\varepsilon_{jt}$ . We stack these return components into  $N_t \times 1$  vectors  $\alpha_t$  and  $\varepsilon_t$  and an  $N_t \times 4$  matrix of betas,  $\beta$ , where  $N_t$  is the number of stocks in existence at time  $t$ . Notice that the individual stock alphas are allowed to vary over time, reflecting that any abnormal returns are likely to be temporary.<sup>32</sup> Following Mamaysky et al. (2008), we assume that stock betas are constant, although this assumption can be relaxed.

Mutual fund returns can be computed by summing across individual stock returns,  $\mathbf{r}_t$ , weighted by the fund's ex-ante portfolio weights at the end of the previous period,  $\omega'_{t-1}$ . Using the decomposition of individual stock returns described above, the excess return of an individual mutual fund ( $i$ ) net of the risk-free rate, transaction costs, fees and expenses ( $k_i$ ) is given by:

$$\begin{aligned} R_{it} &= \omega'_{it-1}(\alpha_t + \beta \mathbf{z}_t + \varepsilon_t) - k_i, \\ &\equiv \alpha_{it} + \beta'_{it} \mathbf{z}_t + \varepsilon_{it}, \end{aligned} \quad (11)$$

where  $\alpha_{it} = \omega'_{it-1} \alpha_t - k_i$ ,  $\beta_{it} = \omega'_{it-1} \beta$ , and  $\varepsilon_{it} = \omega'_{it-1} \varepsilon_t$ .<sup>33</sup>

Mamaysky et al. (2008) represent skills through a persistent, latent process which reflects the fund's ability to process private information.<sup>34</sup> Suppose that fund  $i$  receives a private signal,  $F_{it}$ , which follows an autoregressive process,

$$F_{it} = \nu_i F_{it-1} + \eta_{it} \quad \text{for } \nu_i \in [0; 1). \quad (12)$$

where  $\eta_{it}$  is independently and identically distributed (i.i.d.) with zero mean. The innovations,  $\varepsilon_{it}$ , in (11) and  $\eta_{it}$  in (12) are assumed to be mutually independent and normally distributed. Following Mamaysky et al. (2008), assume that (i) fund portfolio weights are linear in the private signal,  $\omega_{it-1} = \bar{\omega}_i + \gamma_i F_{it-1}$ ; (ii) there is a constant mapping  $\bar{\alpha}_i$  from fund  $i$ 's private signal into its ability to identify abnormal performance,  $\alpha_{it} = \bar{\alpha}_i F_{it-1}$ . These assumptions and equation (11) imply that fund  $i$ 's beta and alpha are linear and quadratic functions of the signal,  $F_{it-1}$ , respectively:

$$\begin{aligned} \alpha_{it} &= \bar{\omega}'_i \bar{\alpha}_i F_{it-1} + \gamma'_i \bar{\alpha}_i F_{it-1}^2 - k_i \\ &\equiv a_i F_{it-1} + b_i F_{it-1}^2 - k_i, \end{aligned} \quad (13)$$

$$\begin{aligned} \beta_{it} &= \bar{\omega}'_i \beta + \gamma'_i \beta F_{it-1} \\ &\equiv \bar{\beta}_i + \mathbf{c}_i F_{it-1}. \end{aligned} \quad (14)$$

---

<sup>32</sup>Liu and Timmermann (2013) develop a theoretical model with temporary abnormal returns (alphas) in the context of convergence trading.

<sup>33</sup>As in Mamaysky et al. (2008) these are assumed to be proportionate to the fund's assets under management.

<sup>34</sup>Kosowski (2011) models manager skills as a latent variable driven by a regime-switching process. As in Mamaysky et al. (2008), this gives rise to a filtering problem, although the filter is non-linear in this case.

The fund manager’s signal,  $F_{it-1}$ , is unobserved to the econometrician, but an estimate of it can be obtained from the fund’s observed returns. To this end, we put the model into state space form:

$$\begin{aligned} R_{it} &= a_i F_{it-1} + b_i F_{it-1}^2 - k_i + (\bar{\beta}'_i + \mathbf{c}'_i F_{it-1}) \mathbf{z}_t + \varepsilon_{it} \\ F_{it} &= \nu_i F_{it-1} + \eta_{it}. \end{aligned} \tag{15}$$

As explained by Mamaysky et al. (2008), the parameters of this model can be estimated fund-by-fund using an extended Kalman Filter that accounts for the presence of the squared value of the state variable,  $F_{it-1}$ , in equation (15).<sup>35</sup>

### 3.1.1. Introducing Information from Fund Holdings

Conventional approaches to measuring mutual fund performance base their inference on time-series estimates using past returns, which can be very noisy. This limits the ability of return-based methods—such as that of Mamaysky et al. (2008)—to identify funds with superior performance.

One way to address this issue is to augmenting the model to account for information from portfolio holdings which, as is clear from equation (13), can be used to track how a fund’s alpha evolves through time. Building on this idea, we next generalize the methodology in Mamaysky et al. (2008) and show that holdings-based information can be added in the form of an additional measurement equation in the state space representation of the model (15). Moreover, this model can be estimated by means of an extended Kalman filter.

Data on fund holdings allow us to perform risk-adjustment at the individual stock level by matching each stock to a portfolio of stocks with similar characteristics in terms of their sensitivity to book-to-market, market capitalization, and price momentum factors. The difference between an individual stock’s return and the return on its characteristics-matched portfolio can be used as a measure of that stock’s abnormal return. Weighting individual stock abnormal returns across all stock positions held by a fund, we obtain the fund-level characteristic-selectivity ( $CS$ ) measure of Daniel, et al. (1997),

$$CS_{it} = \boldsymbol{\omega}'_{it-1} (\mathbf{r}_t - \mathbf{r}_{bt}). \tag{16}$$

Here,  $\mathbf{r}_t$  and  $\mathbf{r}_{bt}$  are vectors of excess returns on stocks ( $\mathbf{r}_t$ ) and benchmark portfolios ( $\mathbf{r}_{bt}$ ), respectively. Benchmark portfolios are chosen to match, as closely as possible, the characteristics of the individual stocks and so  $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{bt}$ . Because the characteristic-matched stocks are chosen mechanically, the average stock can be expected to have zero alpha,  $\boldsymbol{\alpha}_{bt} = \mathbf{0}$ . Using (13) and (16)

---

<sup>35</sup>Normalizing one of the elements of  $\mathbf{c}_i$ , the four-factor model requires 13 parameters to be estimated.

we, therefore, have

$$\begin{aligned}
CS_{it} &= \boldsymbol{\omega}'_{it-1} (\boldsymbol{\alpha}_t + \boldsymbol{\beta}' \mathbf{z}_t + \varepsilon_t - (\boldsymbol{\alpha}_{bt} + \boldsymbol{\beta}'_b \mathbf{z}_t + \varepsilon_{bt})) \\
&= \boldsymbol{\omega}'_{it-1} (\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{bt}) + \boldsymbol{\omega}'_{it-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_b)' \mathbf{z}_t + \boldsymbol{\omega}'_{it-1} (\varepsilon_t - \varepsilon_{bt}) \\
&= \alpha_{it} + k_i - \boldsymbol{\omega}'_{it-1} \boldsymbol{\alpha}_{bt} + (\boldsymbol{\beta}_{it} - \boldsymbol{\beta}_{bt})' \mathbf{z}_t + \varepsilon_{it} - \varepsilon_{bt} \\
&= a_i F_{it-1} + b_i F_{it-1}^2 + \varepsilon_{it} - \varepsilon_{bt}.
\end{aligned} \tag{17}$$

Since the CS measure does not depend on estimated risk factor loadings obtained over some prior historical period, it has the potential to generate a more accurate estimate of fund performance and, thus, to improve inference based on the return-only models.<sup>36</sup>

Our generalized latent skill, holding-based model can be summarized as

$$\begin{aligned}
\begin{pmatrix} R_{it} \\ CS_{it} \end{pmatrix} &= \begin{pmatrix} a_i F_{it-1} + b_i F_{it-1}^2 - k_i \\ a_i F_{it-1} + b_i F_{it-1}^2 \end{pmatrix} + \begin{pmatrix} \bar{\boldsymbol{\beta}}_i + \mathbf{c}_i F_{it-1} \\ 0 \end{pmatrix} \mathbf{z}_t + \begin{pmatrix} \varepsilon_{it} \\ \varepsilon_{it} - \varepsilon_{bt} \end{pmatrix} \\
F_{it} &= \nu_i F_{it-1} + \eta_{it}
\end{aligned} \tag{18}$$

Compared to the model of Mamaysky et al. (2008) in (15), this model uses the additional information contained in  $CS_{it}$ , which has the potential to make estimation and extraction of the fund's private information component,  $F_{it}$ , more precise. We can, again, estimate the parameters of (18) using the extended Kalman filter, now using two measurement equations; Appendix A provides further details.<sup>37</sup>

## 4. Performance Results

This section introduces our data and establishes a performance benchmark based on the popular decile sorting methodology widely used in academic studies. Next, we analyze the empirical performance of portfolios based on funds included in the fund confidence sets.

<sup>36</sup>The CS measurement equation is “idealized,” in the sense that the CS performance measure will not capture intra-quarter trading returns. Note that the CS and return-based measures can be viewed as different estimates of the same underlying fund performance, observed with different measurement errors.

<sup>37</sup>We note an additional important difference between our approach and that of Mamaysky, et al. (2008). Before forming portfolios based on the conditional alpha estimates, Mamaysky, et al. (2008) trim the set of funds. By contrast, in our approach, funds that are eligible for inclusion at a given point in time are assigned to an “active pool”, while excluded funds are assigned to a “passive pool.” Funds can enter the passive pool at any point in time and return to the active pool again. The funds are allocated to the pools following two steps. First, funds whose alpha forecast for the previous month had the same sign as the fund's return in excess of the return on the market portfolio during that month, stay in the active pool for the current period. Second, funds with alpha forecasts less than -200 bps/month or greater than 200 bps/month, or funds whose predicted betas are less than zero or greater than two are moved to the passive pool. Our approach does not require that we assign funds to such active or passive pools.

#### 4.1. Data

Our empirical analysis uses monthly returns on a sample of U.S. equity mutual funds over the 32-year period from 1980:06 to 2012:12. Individual fund returns are taken from the CRSP mutual fund data base, and are net of transaction costs and fees.

To construct our estimate of the  $CS$  measure, we use quarterly holdings data from Thomson Financial.<sup>38</sup> We use these quarterly holdings data to construct a three-month estimate of  $CS$ . We merge the returns and holdings data using the MFLINKS files of Wermers (2000), which have been updated by the Wharton Research Data Services, and allows us to map the Thomson holdings data to CRSP returns using the funds' WFICN identifier. We require each fund to have at least six months of contiguous returns data. In total, we have returns and holdings data on 2,480 actively managed U.S. domiciled equity mutual funds, but we exclude 255 sector funds, and, so, end up with 2,225 funds.<sup>39</sup> The number of funds included in the analysis peaks at above 1,700 in 2009 before declining to around 1,500 in 2012.

For each fund, we obtain an alpha estimate using time-series data on the fund's historical returns. Funds with a very short return record tend to generate noisy alpha estimates. To avoid that our analysis becomes dominated by such funds, we require funds to have a return record of at least five years. Table 2 provides summary statistics for the cross-sectional distribution of individual fund (ex-post) alphas. The median fund has a negative alpha of -74 bps/year. The finding that the median fund underperforms on a risk-adjusted basis is consistent with previous academic studies (e.g., Carhart, 1997). The bottom 5% of funds, ranked by alpha performance, have a negative alpha estimate of about -38 bps/month, or just under -4.5%/year. The top 5% of funds have an alpha estimate of 21 bps/month, approximately 2.5%/year.<sup>40</sup>

#### 4.2. Performance of Decile-sorted Portfolios

To establish a reference point for our FCS results, we first follow the common practice of ranking individual funds' alphas and forming decile portfolios. This approach can be used to see if funds that are expected to have the highest alphas do indeed produce better subsequent performance than lower-ranked funds. Specifically, each month,  $t$ , we rank funds based on their expected alphas  $\hat{\alpha}_{i,t+1|t}$ . We then form ten equal-weighted decile portfolios with the first portfolio (P1) containing the bottom 10% alpha-ranked funds, the next decile containing funds ranked in the second-lowest 10%, continuing up to the top 10% of alpha-ranked funds (P10). To obtain more detailed results

---

<sup>38</sup>Prior to 2004, funds were only required to report holdings every six months. However, the majority of funds disclosed holdings every three months to Thomson. See Wermers (1999) for details.

<sup>39</sup>"Sector funds" are defined as funds whose  $R^2$  is less than 70% in a four-factor regression. For such funds, the simple four-factor risk-adjustment approach is not appropriate, and these funds are, therefore, excluded.

<sup>40</sup>Similar estimates of the cross-section of alphas (not reported here) are obtained for a conventional four-factor model with constant alphas and for the basic model proposed by Mamaysky et al. (2008).

for the bottom and top funds, following Carhart (1997), we also divide P1 into the bottom 5% alpha-ranked funds and funds ranked between the bottom 5% and 10% (labeled P1A and P1B, respectively); we use a similar split for the P10 portfolio (P10A and P10B). Finally, we record the returns on each of these portfolios over the subsequent month. Each month, as new data arrive, we repeat this sorting routine, form equal-weighted portfolios based on the funds' updated alpha estimates, then, again, record their next-month returns. We use five years of data to initiate the portfolio sorts, and another five years to obtain an estimate of the predictive alpha and, thus, generate a time series of portfolio excess returns,  $R_{pt}$ , over the 21-year period from July 1, 1990 to December 31, 2012. This approach, while requiring 10 years of returns history for each fund under consideration, does not induce substantial survival bias, as we do not require funds to survive beyond a single month into the future.<sup>41</sup>

To evaluate the performance of the portfolios, we follow conventional practice and estimate four-factor alphas on the (out-of-sample) portfolio returns,

$$R_{pt} = \alpha_p + \beta_p' \mathbf{z}_t + \varepsilon_{pt}, \quad t = 1, \dots, T. \quad (19)$$

The resulting estimates,  $\hat{\alpha}_p$ , can be interpreted as the “average” portfolio alphas.

Table 3 presents alpha estimates for the decile portfolios. We find strong evidence of negative and statistically significant alpha estimates for the bottom five ranked decile portfolios (P1-P5). The underperformance of these decile portfolios ranges from -9 bps/month to -16 bps/month, and is qualitatively similar across the three different models used to rank funds. The alpha estimates of the top-ranked decile portfolio (P10) is 2 bps/month and is statistically insignificant, indicating that the conventional portfolio sorting approach fails to identify funds with abnormal positive performance. This conclusion carries over from the top 10% to the top 5% portfolio of funds (P10B), which only performs marginally better than the P10 portfolio.

Despite these shortcomings, the portfolio sorting approach does succeed in differentiating between the best and worst performing funds as the estimated differential in alphas between the P10 and the P1 portfolios is large and positive (18 bps/month) and highly statistically significant. This finding is, however, driven mostly by the identification of funds with inferior performance. As an alternative test of the value of the ranking information in the portfolio sorts, the last line in Table 3 reports the MR test for a monotonic pattern in the alphas proposed by Patton and Timmermann (2010). The null is that there is a flat or declining pattern in the alphas, while the alternative is that there is a monotonically increasing pattern as we move from  $P1$  to  $P10$ . Small  $p$ -values for this test are evidence that a model succeeds in ranking the future risk-adjusted performance of the funds. The test statistic is highly statistically significant, suggesting that the performance model contains valuable ranking information despite its failure to identify funds with large positive

---

<sup>41</sup>See also Carhart (1997) for a discussion of survivorship bias.

alphas.

We conclude from these findings that the conventional portfolio sorting approach can be used to identify a broad set of funds with inferior performance, but is not well suited for identifying funds with significantly superior future performance.

#### 4.3. Performance of Superior FCS Funds

We next analyze the performance of our FCS approach, which forms portfolios from the set of funds identified as top performers. Figure 1 (top panel) shows how the FCS approach selects funds that stand out even among funds with positive forecasts of alpha. The figure plots the distribution of predictive alpha estimates obtained using our performance model at a single point in time (July 2006). The black curve shows the distribution of predictive alphas for the full set of funds in existence at that point, i.e., the population of funds. This curve is centered to the left of zero, and has a wide dispersion. We also show the distribution of predictive alphas for funds with positive predicted alphas (green), and funds whose alpha estimates are positive and statistically significant using the Romano-Wolf approach (red). Finally, the blue line captures the distribution of predictive alphas for funds included in the FCS. Note that the FCS curve lies to the right of the distribution curve for the more inclusive set of funds with significantly positive predictive alphas, highlighting that the FCS methodology is far more discriminating than an approach that simply selects funds with (significantly) positive alphas.

Turning to the formal analysis, the left column in Table 4 presents alpha estimates for a set of portfolios formed by equal-weighting the funds selected each month by applying the FCS approach to the set of funds with significantly positive alpha, identified using the approach of Romano and Wolf so as to reduce the dimensionality of the second-stage FCS pairwise comparisons. As in Table 3, the (out-of-sample) period covers 1990:07 to 2012:12. We consider a range of values for the significance level of the test,  $\lambda = \{0.10, 0.50, 0.75, 0.90\}$ , corresponding to a wide, medium, medium-tight, and tight FCS. Results for the simple (Carhart, 1997) alpha-ranked top decile of funds (P10) are shown for comparison, and serve as a benchmark given the widespread use of this approach. We focus our discussion on equal-weighted portfolios, but the results also hold for value-weighted portfolios.

We observe a monotonically increasing pattern in alpha performance, rising from 13 bps/month for the wide FCS to 48 bps/month for the tight FCS as we move from  $\lambda = 0.10$  to  $\lambda = 0.90$ . Moreover, the performance of the tight FCS is highly statistically significant, with a t-statistic of 2.36. Compared to the simple Carhart (1997) approach, the tight FCS portfolio produces an outperformance of 46 bps/month.

Our FCS approach is novel, and so it is important to understand the mechanisms that are at play. The most obvious effect comes from the breadth of the funds eliminated by the FCS

approach. As we reduce  $\lambda$  from 0.90 to 0.75, 0.50, or 0.10, and, in doing so, move from the tight to the medium-tight, medium, then wide set of funds, fewer funds are eliminated and the performance of the FCS portfolios is markedly reduced. This suggests that the additionally included funds tend to perform worse than the funds identified by the more selective approach. This finding is unsurprising, because many funds have alphas whose estimates are surrounded by large standard errors. For these funds, choosing a small value of  $\lambda$  means that the equivalence test used to eliminate inferior funds has insufficient power to reject the null that the fund alphas are identical; we analyze this finding in more detail below.

Second, the second-step sequential elimination of funds deemed to have statistically significant alphas in the first step, is important to the FCS results. In particular, the alpha estimate of a portfolio comprising the initial set of funds identified by the Romano-Wolf procedure is 16 bps/month—very similar to the performance of the wide FCS portfolio, but much smaller than that of the tight FCS portfolio.

Third, using informative conditioning information to predict the future performance of funds is important to the FCS approach. When applied to a simple, constant-alpha four-factor model, we do see an improvement in the alpha performance of the FCS portfolio as we move from the FCS-wide to the FCS-tight portfolio, and we also see an improvement over the portfolio of top-decile funds. However, these improvements are much smaller in magnitude than what we observe for our more refined performance measurement model. A similar conclusion holds when the FCS approach is applied to the approach of Mamaysky et al. (2008) that does not use holdings information. Interestingly, however, adding holdings information to the Mamaysky et al. (2008) model does not, on its own, increase average performance either for funds in the top decile or for funds with significantly positive alpha estimates, as identified by the Romano-Wolf approach.

These observations show that it is the joint effect of using more informative performance signals (pooling return and holdings data to obtain more accurate alpha estimates) and using a fund selection approach with stronger power to discriminate between funds (i.e., eliminating inferior funds by means of the FCS approach) that produces substantially improved risk-adjusted return performance from our fund selection approach.

#### 4.4. Performance of Inferior FCS Funds

Next, we analyze whether the FCS approach can be used to identify funds with inferior performance. To this end we need to redefine the objective function. We do so by maximizing a “predictive alpha” loss function which focuses on funds in the left tail of the performance distribution:

$$L_{i,t} = \text{Min}(\hat{\alpha}_{i,t|t-1}, 0) \text{sign}(R_{i,t} - \beta'_{i,t} \mathbf{z}_t). \quad (20)$$



Recall that the sign function  $\text{sign}(\bullet)$  equals -1 if its argument is negative. Hence,  $L_{i,t}$  will be large for funds expected to perform poorly (through  $\text{Min}(\hat{\alpha}_{i,t|t-1}, 0)$ ) with negative risk-adjusted returns, i.e.,  $\text{sign}(R_{i,t} - \beta'_{i,t}\mathbf{z}_t) = -1$ . Funds with zero or positive expected alphas ( $\hat{\alpha}_{i,t|t-1} > 0$ ) get excluded from consideration since  $L_{i,t} = 0$  for such funds.

Again, we base our analysis on a sample estimate of the mean performance for fund  $i$ ,

$$\bar{L}_{i,t} = \frac{1}{t - t_{i0}} \sum_{\tau=t_{i0}+1}^t \text{Min}(\hat{\alpha}_{i,\tau|\tau-1}, 0) \times \text{sign}(R_{i,\tau} - \hat{\beta}'_{i,\tau}\mathbf{z}_\tau). \quad (21)$$

As a first illustration of the inferior funds selected by the FCS approach, the bottom panel of Figure 1 shows the distribution of predictive alphas for the funds that were in existence (and had a minimum of five years of data) on July 1st, 2006. The distribution of predictive alphas for funds in the inferior FCS (blue line) is centered around -0.5%/year, and is much further to the left than the distribution curve for funds with negative alpha estimates (green line), funds with significantly negative alphas (red line), or for the population of funds as a whole (black curve). As in the case of funds with superior performance (top panel), the FCS approach is more selective in identifying funds with inferior performance than an approach that simply considers funds with negative expected performance.

The right column in Table 4 considers the performance of a strategy that first uses the bootstrap method of Romano and Wolf (2005) to identify funds with significantly negative alphas, followed by the use of the FCS approach to select the set of worst performers among these funds. While the tight FCS portfolio generates the lowest alpha estimate, there is no longer a monotonic relation between tightness ( $\lambda$ ) and the magnitude of the alpha estimate. Even so, at -23 bps/month, the alpha estimate of the tight FCS portfolio is more negative than those of both the decile of worst performing funds (-16 bps/month) and the portfolio of funds with significantly negative alphas as identified by the Romano and Wolf (2005) approach (-16 bps/month). Similar results are obtained when we use other performance benchmarks such as the constant-alpha, four-factor model of Carhart (1997) or the Mamaysky et al. (2008) model that does not use holdings information. The reason for the low sensitivity to the value of  $\lambda$  found for the worst-performing funds is easy to understand. As can be seen from Table 2, the left tail of the performance distribution is much thicker than the right tail, and there is also greater persistence among inferior funds than among superior funds. This means that varying  $\lambda$  has less of an impact on the results for the inferior funds than its effect on the set of superior funds.

We conclude the following from these results. First, the FCS approach is capable of selecting funds whose performance is far better than that achieved by funds included in the conventional (e.g., Carhart, 1997) rank-based top-decile portfolio. Second, the performance of the FCS approach is best when we allow the approach to eliminate more funds and become more selective. This is

consistent with our finding from the analysis of the decile portfolio’s performance that the ability to outperform is not very widespread among the funds in our sample.

#### *4.5. Time Variation in Fund Performance*

The alpha estimates reported in Tables 3 and 4 show the average performance of a set of decile-sorted and FCS portfolios, respectively. Such estimates do not reveal if the performance of different funds is concentrated in certain states or occur over certain periods of time. Addressing this point can help provide us with further insights into the nature of the performance of the selected funds. Accordingly, we next explore this issue using a non-parametric methodology designed to capture “local” time variation in portfolio alphas. Appendix B provides further details of this approach.

Figure 2 shows the evolution through time in the risk-adjusted performance (alpha) of superior funds included in the FCS portfolio (in blue), along with the performance of the traditional top decile portfolio (P10, in purple). The red line tracks the performance of the initial set of funds, selected using the Romano Wolf first-stage filter. Finally, the green line tracks the performance of the set of funds with positive predictive alphas, while the black line tracks the performance of a portfolio containing all funds in existence at a given point in time.

From 1993-1995, the local alpha estimates of the FCS portfolio are either close to zero or positive. Starting in 1996, the performance of the FCS portfolio increases dramatically, before peaking at about 200 bps/month in 1999-2001. Over the same five-year period the performance of the P10 decile portfolio hovers around zero. The FCS portfolio only performs poorly during brief spells in 1990 and around 2004. After 2007, the FCS portfolio produces positive alphas of roughly 25-50 bps/month.

Figure 3 plots the time-series evolution in the local alpha estimates for the inferior FCS portfolio. For this portfolio, we see substantially worse performance during the first five years of the sample followed by slightly better performance during 1998-2002 and, again, in 2012.

#### *4.6. Technology Factor*

Our results up to this point use a four-factor model to compute the fund-level alpha forecasts, compare fund performances using the FCS methodology, and, finally, evaluate fund performance. One concern that is commonly raised when assessing results based on factor models is that they could be affected by an omitted risk factor. As we show in Section 5.1, there are times when the FCS of superior funds is quite concentrated and loads up on technology-heavy funds, so a concern here might be that our results are not robust to the presence of a technology factor.

To address this issue, and as a robustness check, we therefore include a fifth, technology index, factor in all three steps described above. Our performance model already requires estimation of 13 parameters and the inclusion of an additional factor increases this number to 15. This poses a

serious challenge to obtaining reliable parameter estimates and so we expand the estimation window for the five-factor model from 60 to 120 months. Adding on the five years of data used to conduct pair-wise fund comparisons, the out-of-sample evaluation period is cut short to 1995:7-2012:12—a sample that excludes the period with the highest performance of the FCS approach (see Figure 2). Funds selected by a relatively tight FCS ( $\lambda = 0.8$ ) in this case go on to produce five-factor alpha estimates of 18 bps/month with a  $t$ -stat of 2.16—a similar  $t$ -stat as that obtained (2.47) for the four-factor performance model applied to the shorter sample.

In an additional robustness exercise, we compute the non-parametric characteristics selectivity (CS) performance measure for the FCS portfolio that uses the five factor specification. The average value of the CS measure is 15 bps/month with an associated  $p$ -value of 0.000.

These results suggest that including a technology factor has a relatively modest effect on our results.

#### 4.7. Fund Performance in High and Low Volatility States

Kacperczyk et al. (2016) analyze mutual funds' allocation of attention in a setting with multiple risky assets. Provided that the uncertainty about the aggregate risk factors is sufficiently large, their analysis shows that an increase in the variance of the aggregate risk factors (i) increases the dispersion in fund portfolio holdings; (ii) increases the dispersion in fund excess returns (Proposition 3 in Kacperczyk et al. (2016)); and (iii) increases the excess returns for skilled funds (Proposition 5). The intuition underpinning these results is that rewards from processing information are higher during times when asset payoffs are highly volatile, leading to stronger incentives for skilled investors to allocate extra attention to their portfolio decisions during such times. Conversely, unskilled investors can either not increase their attention (noise traders) or face higher costs from doing so, with the result that the differential between skilled and unskilled funds' return performance increases during periods of high volatility.

To test the first prediction (wider dispersion in portfolio holdings and returns), we compute the cross-sectional dispersion in industry exposure across the funds in our sample. At any point in time, the dispersion measure is computed relative to the average industry exposure across all funds. We then sort the sample into three groups, according to whether the aggregate (market) volatility is low, medium or high, assigning one-third of the total sample to each volatility bin.<sup>42</sup>

Panel A in Table 5 (left column) shows that the dispersion in industry concentration—computed using holdings in 48 industries—across funds is twice as high during periods with high aggregate risk compared to periods with low aggregate risk (1.023 vs. 0.464).<sup>43</sup> Moreover, the difference in

<sup>42</sup>Our empirical tests of the relation between fund performance and aggregate market volatility all use a “local” measure of volatility, created using a 12-month rolling window volatility estimate and based on daily returns.

<sup>43</sup>1.023 corresponds to an average deviation of an industry's weight in a given fund of about 14%, relative to the industry's average portfolio weight across funds, while 0.464 corresponds to a deviation of only about 10%.

mean industry concentration across states with high and low aggregate volatility is statistically significant with a  $p$ -value of 0.023.

Next, consider the second prediction implied by the analysis in Kacperczyk et al. (2016), namely that excess returns should be more dispersed across funds during times of high volatility. Consistent with the prediction, the right column of Panel A in Table 5 shows that during periods with high aggregate risk the dispersion in the CS measure is two and a half times greater (6.09 versus 2.29) than during periods with low aggregate risk. Once again, this difference in dispersion across the high and low volatility states is statistically significant. Since the CS measure is constructed from funds' individual security holdings, this finding is consistent with the cross-sectional dispersion increasing during times of high volatility not simply due to greater differences in fund industry allocations, but due to more dispersion in individual stock holdings.

Our FCS approach is ideally suited for testing the third prediction—that excess returns should be higher among the most skilled funds during times with high volatility—as it seeks to identify the set of funds with superior performance and thus, as we have seen, provides a direct estimate of the breadth of skill among mutual funds at a given point in time. Even so, it can be challenging to identify, *ex-ante*, outperforming active equity funds during periods of high stock market volatility. Although the greater separation in fund alphas implied by the theory should enhance our ability to identify superior funds, test procedures are hindered by a larger level of noise in each point estimate during periods with high volatility. Which of these two influences dominates depends on how much the skilled fund managers' performance increases during periods with higher volatility.

Panel B1 in Table 5 reports four-factor alpha estimates computed for the low, medium and high risk samples, using a 12-month rolling window to determine the underlying volatility state. The FCS alphas are notably higher (64 bps per month) during periods with high aggregate volatility (left columns) than during periods with low volatility (essentially zero bp/month), and their difference is highly statistically significant. An even stronger result holds when we sort by idiosyncratic volatility (right columns), where the alpha of the FCS portfolio is slightly negative during low-risk episodes but 79 bps/month during periods with high idiosyncratic risk. Conversely, the portfolio of the top-10 percent of fund managers (P10) does not detect this relation between volatility and performance—in large part because it fails to account for the effect of high volatility on the breadth of performance skills.

Similar, if even stronger, results hold when we use a 36-month rolling window to determine the underlying volatility state. Using a 36-month window produces a less noisy estimate of the underlying volatility state, although it may also fail to detect short-lived durations of unusually high or low volatility. Using the longer 36-month window (Panel B2), we find that performance increases monotonically from 12 bps/month in the low aggregate risk state to 82 bps/month in the high (aggregate) volatility state or from 7 bps/month to 104 bps/month when sorting on idiosyncratic volatility. P10 detects a relatively small, but significant difference between low and

high volatility periods when using a 36-month rolling window.

We conclude from this evidence that performance among superior funds is notably higher during periods with high aggregate or idiosyncratic volatility and that this effect outweighs any increased noise that is introduced to estimated alphas during such periods. The FCS efficiently locates the set of outperforming active equity funds during the periods when they produce their greatest levels of alpha. In contrast, the conventional decile-sorting approach falls notably short at identifying the improved performance in high-volatility states.<sup>44</sup>

## 5. Selection of Superior and Inferior Funds

The previous section shows that the FCS approach can be used to successfully identify funds with superior or inferior risk-adjusted performance. The FCS approach is fundamentally different from existing methods, such as decile sorting which keeps the proportion of “top” funds fixed through time. In contrast, the FCS approach endogenously determines how many funds to include. This section provides details of both the number of funds selected by the FCS approach and their identify.

### 5.1. Identifying Superior Funds

We can gain important economic insights by studying how the set of superior funds evolves over time. Specifically, the turnover in the set of funds identified to be superior is related to the skill level and duration of individual funds’ ability to outperform and the speed with which the competition catches up.

The top panel of Figure 4 shows the number of funds included in the FCS portfolio at each point in time. Because this set depends on the size of the test,  $\lambda$ , we present results for the tight, medium-tight, medium, and wide sets,  $\lambda = \{0.9, 0.75, 0.5, 0.1\}$ , but focus on the tight set ( $\lambda = 0.90$ ) in most of our discussion. The number of funds included in the top FCS fluctuates considerably over time. Around 5-10 funds get selected during the two-year period from 1992-1994, followed by a narrower FCS up to 2000. As shown in Figure 2, this period coincides with the best risk-adjusted performance of the FCS portfolio, suggesting that, at least for this historical period, our methodology succeeds in identifying a narrow set of funds with high subsequent performance. After 2000, the FCS undergoes considerable change with a broad array of funds entering and leaving the FCS up to 2007. During the last 18 months of the sample, the top FCS grows in size, and includes around 20 funds. On average, the tight FCS ( $\lambda = 0.90$ ) includes 8 funds, while the medium-tight

---

<sup>44</sup>Kacperczyk et al. (2016) compare fund performance during recession and expansion periods, noting that recessions tend to be periods with higher stock market volatility. We also considered splitting the data sample using the NBER recession indicator, but found that fund performance was imprecisely estimated, partly due to the small number of recession months in our sample (34), which made the estimated recession alpha imprecise. Instead we find that our volatility measure provides a better proxy for high versus low volatility periods.

( $\lambda = 0.75$ ), medium ( $\lambda = 0.50$ ) and wide ( $\lambda = 0.10$ ) FCS include 12, 17 and 25 funds, respectively.<sup>45</sup>

Because the number of funds in our data increases over time, it is useful to also consider the percentage of funds identified by the FCS approach as being top performers. The bottom panel in Figure 4 provides a time-series plot of this proportion for the three different values of  $\lambda$  considered above. The percentage of top performers varies greatly over time. It fluctuates between 0.5% and 4% in the early part of the sample (assuming  $\lambda = 0.90$ ), and is quite low up to 2000. After this period, 2% or more of the funds regularly get selected, and the proportion of included funds peaks at around 3% in 2011-2012. For smaller values of  $\lambda$ , the percentage of funds included is, of course, a bit higher, but it rarely exceeds 6%. On average, the tight FCS includes 1.8% of the funds, while the medium-tight, medium and wide FCS portfolios include 2.9, 4.1 and 5.8% of the funds, respectively.

Figure 5 presents more detail on the individual funds identified as superior by the FCS approach. Specifically, for each fund that gets selected at least once during the sample, the figure shows when this particular fund is chosen. (The labeling on the y-axis is arbitrary, but maps one-to-one to the fund ID, and is sorted chronologically according to the first appearance of a particular fund in the FCS.) The top panel of Figure 5 presents results for the medium-tight FCS, while the bottom panel assumes that  $\lambda = 0.9$ . For both values of  $\lambda$ , a single fund—Fidelity Select Technology—gets selected for most of the five-year period from 1995-2000. Between 2008 and 2011, a single fund—T Rowe Price Media and Telecommunications—is again selected. For most of the remaining periods a broader array of funds is selected, consistent with Figure 4.<sup>46</sup>

To interpret these findings, note that the number of funds identified as being superior at a given point in time depends on the ability of the underlying performance model to accurately estimate individual funds' risk-adjusted returns and, thus, discriminate between funds. For example, during periods where individual fund alphas are estimated with large sampling errors, we would not expect  $t$ -tests such as equation (7) to be able to eliminate many funds from the FCS which, thus, might include many funds. In contrast, during times where individual funds' track records are more informative and alpha estimates are sharper, we would expect the FCS approach to be more discriminating, resulting in a narrower set of funds being included.

Figure 6 illustrates this point by showing the predictive alpha of the fund(s) included in the FCS (marked in blue) at three points in time (January 1997, 2002, and 2010) along with the alphas of those funds excluded (marked in red). In the top and bottom plots a single fund gets selected by the FCS due to its markedly superior performance. In contrast, no single fund stands out in

---

<sup>45</sup>Recall that the  $\lambda$  parameter should be chosen to reflect the trade-off between an investor's utility from correctly including a genuinely superior fund versus her disutility from wrongly including an inferior fund.

<sup>46</sup>The ability of the FCS to identify funds with superior performance can also be explored by studying the average value of the  $t$ -test used to eliminate inferior funds through time. This value hovers between -2 and -6 through most of our sample. Periods when the  $t$ -test is less negative (closer to zero) are times when a broader set of funds have more similar levels of expected performance; periods with more negative  $t$ -tests coincide with a narrow set of funds that are expected to outperform (and, thus, other funds are easy to eliminate through the sequence of  $t$ -tests).

the middle panel, giving rise to a wider fund confidence set.

We conclude the following from our analysis. First, our ability to identify superior funds fluctuates significantly over time—sometimes this set is quite broad, containing up to 5% of the funds, at other times, the set is very narrow and contains less than a handful of funds—or even just a single fund. Second, the set of superior funds is almost always much smaller than the 10% figure assumed in decile ranking studies, and has been decreasing over time. This means that a procedure that forms portfolios based on a fixed fraction of the total number of funds—such as the conventional decile sorting approach—is likely to grossly underestimate the possibility of identifying funds with superior performance.

### 5.2. *Identifying Inferior Funds*

Figures 7 and 8 present plots similar to those presented in figures 4 and 5, but now applied to the set of funds identified to have inferior performance. Figure 7 shows that the set of funds with inferior performance fluctuates considerably through time, peaking at close to 5% for the narrow FCS portfolio ( $\lambda = 0.90$ ), but often containing 1% or fewer of the funds. On average, the tight FCS ( $\lambda = 0.90$ ) includes 15 funds, while the medium-tight ( $\lambda = 0.75$ ), medium ( $\lambda = 0.50$ ) and wide FCS ( $\lambda = 0.10$ ) portfolios on average include 24, 35 and 59 funds, respectively. These figures correspond to 3.0%, 4.8%, 7.0% and 11.6% of the total number of funds, respectively. Thus, across all values of  $\lambda$ , the set of inferior funds is broader and includes more funds than the equivalent FCS for superior funds.

Figure 8 shows the inclusion plot for individual funds with inferior performance. The top panel assumes that  $\lambda = 0.75$ , while the bottom panel assumes  $\lambda = 0.9$ . We continue to see periods during which only a single inferior fund is selected by both the tight, and medium-tight FCS approach. For example, between 1993 and 1995, only the Centurion Growth Fund gets included in the tight FCS, while, between 2001 and 2004, the tight FCS of inferior funds predominantly consists of the Phoenix Oakhurst Strategy F fund.

Interestingly, passive index funds, such as the Vanguard S&P500, are never included in either the superior or inferior FCS portfolios, suggesting that, at each point in time, there always exist funds with genuinely better or genuinely worse ex-ante expected performance than a passive low-cost strategy.

An alternative way to address how difficult it is for mutual funds to generate abnormal performance and how this varies through time is by studying how many funds have significantly positive or negative alpha estimates. The top panel in Figure 9 plots smoothed kernel estimates of the proportion of funds with a statistically significant alpha estimate, identified using the methodology of Romano and Wolf (2005), which is ideally suited to address this question. The number starts at 12% in 1990 (approximately 25 funds), drops to less than 5% in 1998 (approximately 15-20 funds),

and ranges between 5% and 8% for the remainder of the sample, corresponding to between 40 and 75 funds. This shows that there is significant time variation in the scope for individual funds to outperform. The downward trend from 1990 through 1997 and, again, from 2002 until the end of the sample (2012) is consistent with competition intensifying through time (Pastor, et al. (2015)).

The bottom panel of Figure 9 plots the proportion of funds with significantly negative alpha estimates. On average, about 20-25% of funds have significantly negative alpha estimates, however, the scope for underperformance varies substantially through time. Fewer than 15% of funds had significantly negative alphas in 1994 and from 2010-2011, whereas more than 25% of funds had significantly negative alpha estimates during much of the decade from 1999-2009. At the end of the sample, three times as many funds had significantly negative alphas as funds with significantly positive alphas.

### 5.3. Turnover and Duration of Funds in FCS

As a way to illustrate how often individual funds enter and exit the FCS, we next compute the turnover in the set of funds included in the tight FCS portfolio, and compare this to the turnover among funds in the top and bottom decile portfolios. Specifically, we compute

$$Turnover_t = \frac{1}{2} \sum_{i=1}^{N_t} |\Delta w_{it}|, \quad (22)$$

where  $w_{it}$  is the portfolio weight on fund  $i$  at time  $t$ , and  $\Delta w_{it} = w_{it} - w_{it-1}$  measures the change in the portfolio weight on fund  $i$  from month  $t - 1$  to month  $t$ .

The average monthly turnover for the top FCS portfolio is 0.23, while the corresponding figure for the bottom FCS portfolio is 0.24. For comparison, the monthly turnover of the top and bottom decile portfolios are 0.12 and 0.11, respectively. Thus, the turnover is somewhat higher for the FCS portfolios, which is what we would expect, given that these portfolios typically contain fewer funds than the more broadly diversified decile portfolios.

To help explain what generates the higher turnover in the FCS portfolios, Figure 10 provides details on the average duration of funds in the FCS (left column) and decile portfolios (right column) with top performers in the upper panels and bottom performers in the lower panels. Focusing on top-ranked funds, the figure shows that the main reason for the high turnover in the top FCS portfolios is that approximately 40% of all funds remain in the FCS portfolio for only one or two months, while another 20% of the funds get selected for three or four consecutive months. In comparison, about 45% of funds remain in the top decile portfolio for four months or less. There are also fewer funds with very long durations in the top FCS portfolio, compared to the top decile portfolio. Similar conclusions hold for the portfolios of bottom-ranked funds shown in the lower panels of Figure 10. These short durations are in part a result of estimation errors affecting



individual funds' alpha estimates but may also be indicative of strong competition which makes it more difficult for individual funds to be repeatedly included in the set of superior funds.

#### *5.4. Business Cycle Variation in the Set of Superior and Inferior Funds*

Our empirical analysis up to this point shows that the width of the set of funds identified to have superior performance varies a great deal over time. To see whether the proportion of funds identified to have either superior or inferior performance depends on the state of the economy, we next perform a simple analysis that regresses the percentage of funds in either the top or the bottom FCS on a range of state variables commonly used to capture the state of the economy. Specifically, we use the 1-month T-bill rate, the term spread, the default spread (i.e., the difference between the yield on portfolios of (Moody's) Baa- and Aaa-rated bonds) and the dividend yield as our state variables. These variables feature prominently in the literature on return predictability for the broad stock market (Welch and Goyal, 2008), and have also been used to model state-dependence in mutual fund performance (Ferson and Schadt, 1996; Avramov and Wermers, 2006; and Banegas, et al., 2013).

Table 6 reports the outcome of this analysis. We find that many of the state variables are significantly correlated with the width of the top FCS portfolio. Notably, the proportion of funds included in the top FCS tends to be higher in environments with a high short interest rate, a steep yield curve (high term spread), and a small default yield.<sup>47</sup>

Turning to the funds in the bottom FCS, we find somewhat different results: the proportion of funds whose performance is judged to be inferior tends to be higher in economic states with high interest rates, a high default spread, and a high dividend yield. A higher default spread is often associated with economic recessions, and, so, this suggests a very different pattern than that uncovered for the FCS comprising funds with superior performance.

We conclude, from this evidence, that economic states with higher expected future growth (higher term spread) and lower risk premia (smaller default yield) are associated with a broader set of funds being identified as superior. Such states are more likely to occur in economic expansions, and, so, our finding is consistent with that of Kacperczyk et al. (2014) that some funds can pick stocks during economic upturns. Conversely, inferior performance tends to become more widespread during recessions.

---

<sup>47</sup>A similar set of results apply if we focus on the number of funds (rather than the proportion of funds), although, now, the T-bill rate is insignificant.

## 6. Style and Industry Exposures

To gain insights into the investment strategies underlying the abnormal returns generated by the FCS portfolios, we next analyze the evolution in the FCS-chosen funds' style exposures. To this end, we exploit that we have access to quarterly holdings for all funds included in our analysis. The holdings data include information on funds' exposure to size, book-to-market, and momentum risk factors, along with their industry concentrations. For each FCS portfolio, we use these data to construct equal-weighted cross sectional averages of the characteristics and industry concentrations for the underlying funds. Because the time series of holdings data can be quite volatile, we consider rolling one-year averages of the portfolio characteristics.

### 6.1. Style Exposures

Table 7 presents attributes of the funds in the top and bottom FCS portfolios and, for comparison purposes, the average computed across all funds. The average attributes are calculated by first taking cross-sectional averages of the relevant set of funds at a given point in time, then averaging this cross-sectional mean over time. The attributes include the characteristic selectivity measure computed over the past one-month (CS-1m) and three-month (CS-3m) horizons, total net assets measured in millions of dollars (TNA), portfolio-holdings-based style attributes for size (Size), book-to-market (BTM), and momentum (MOM) style factors, the gap between actual returns and the hypothetical return based on reported portfolio holdings (return gap), fund flows, net cash inflows, expense ratio, and turnover.

We observe substantial differences in the attributes of superior and inferior funds, and between either of these types of funds and the average fund. Specifically, superior funds have substantially higher CS measures than average funds, which, in turn, have higher CS measures than inferior funds. This is to be expected, since the CS measure is used to estimate fund performance and, thus, determines which funds get allocated to the superior and inferior portfolios.

Interestingly, the portfolio of superior funds consists of funds whose average size (\$3.8bn) is considerably larger than that of both the average fund (\$1.8bn) and inferior funds (\$0.6bn), indicating that the average superior fund performs well enough to more than offset any diseconomies-of-scale. Turning to the style factors, the portfolio of superior funds holds firms with largely similar size, book-to-market ratios, and momentum, compared to the average fund, while the portfolio of inferior funds exhibits some differentiation on the size dimension. Both superior and inferior funds tend to exhibit a similar return gap, which indicates that the return gap is not a useful proxy for the skills captured by the FCS technique.

Further, inflows are much larger for superior funds, as expected (due to their higher past returns). Superior funds also have a much lower expense ratio and somewhat lower turnover than

inferior funds. This is consistent with the findings of Carhart (1997) that higher expense ratios and higher turnover tend to reduce net return performance.

The above-mentioned style exposure results, which were averaged over time, mask important differences during particular time periods. Accordingly, Figure 11 plots the time-series of style exposures for funds in the top (left column) and bottom FCS (right column), along with style exposures for the average fund. Comparing the plots for the top FCS portfolio to that of the average fund, superior funds are more volatile in all three style dimensions. This result indicates that the FCS rotates among funds residing in different styles, depending on macroeconomic conditions.

Funds in the superior FCS portfolio tend to have a heavier allocation to small- and mid-cap stocks than the average fund—the most notable exception occurring during the global financial crisis (2008-2009) where the superior funds overweighted large-cap stocks. The top FCS portfolio tilts towards growth stocks from 1994 to 2001 (consistent with growth stocks having the most uncertainty in valuations during this period), but overweightes value stocks from 2001-2007. As shown in the bottom panel of Figure 11, the top FCS funds have a notably higher exposure to the momentum factor than the average fund only for brief spells during 1998, 2009 and 2012, suggesting that this portfolio is not predominantly capturing rewards for momentum risk.

A very different picture emerges for the inferior FCS funds. These funds overweight large cap stocks between 1993 and 1997, and again after 2009, overweighting small stocks during the interim period from 1997-2008. The inferior FCS funds overweight value stocks for most of the sample, while their exposure to the momentum risk factor only deviates substantially from that of the average fund during 1994-1997.

Figure 12 shows histograms depicting the performance of the top FCS funds (top row) for different realizations of the size, book-to-market, and momentum risk factors sorted into quintiles. For example, the left-most blocks (within each style/FCS segment, e.g., “Size Superior FCS”) show the average risk-adjusted performance of the FCS funds during months with the lowest quintile of realized returns of the different risk factors, while the right-most blocks show the average performance during months in which the realized value of the factors were in the top quintile.

The top left panel reveals a systematic and monotonically increasing relation between the performance of the superior FCS funds and the size factor: the better large-cap stocks perform relative to small stocks, the better funds in the superior FCS perform during a particular month. The relation between the book-to-market factor and the performance of the superior FCS funds shows that superior funds perform far better (with an alpha close to 100 bps/month) during months where this factor was at its lowest, i.e., during months where growth stocks outperform value stocks. Finally, as shown in the right panel, there is no monotonic relation between the performance of the superior FCS funds and the realized value of the momentum risk factor.

The bottom row of panels in Figure 12 presents a similar set of plots for the inferior FCS funds. Inferior FCS funds perform worse in periods where large stocks outperform small stocks (left panel),

when value stocks outperform growth stocks by a large margin (middle panel), and during times with a low realized value of the momentum factor (bottom right panel).

Overall, these results indicate that the alphas of superior and inferior funds have a common component that varies with the performance of the market capitalization and value/growth dimensions in the marketplace and, to a lesser extent, with the outcome of the momentum risk factor.

## 6.2. *Industry Concentration*

Table 8 reports time-series average industry concentrations for the top FCS funds, bottom FCS funds, and for all funds (equal-weighted averages) based on an industry classification that uses 48 industries (obtained from Ken French’s data site). Industry concentrations are also reported separately for high-volatility and low-volatility states. A result that stands out is that funds in the top FCS have much greater average concentrations in communications, business services, computers, and electronic equipment, than funds in the bottom FCS, indicating the importance of manager skill in choosing among technology firms during the 1990 to 2012 period. Moreover, these differences are particularly high in the high-volatility state.

One way to summarize the extent to which the superior and inferior FCS portfolios take industry bets is by calculating “active share” (Cremers and Petajisto, 2009) at the industry level, i.e., one-half times the absolute deviation of each industry weight, measured relative to the average mutual fund’s industry weighting, summed across industries, then averaged across time. For superior FCS funds, this industry active share measure is 27.7 overall versus an active share of 10.8 for inferior FCS funds. Industry active share is twice as high in recessions (45.1 and 21.7 for superior and inferior funds, respectively) compared to expansions (25.7 and 10.1), suggesting that both superior and inferior funds deviate substantially more from the average fund’s industry weights during recessions. Moreover, for the superior FCS funds, industry active share is notably higher in high-volatility states (32.5) than in low-volatility states (25.9), whereas it varies less across these states for the inferior FCS funds.

To get a sense of the dynamic industry tilting strategies implied by the portfolios of superior and inferior funds, Figure 13 plots one-year rolling averages of industry concentrations for the top FCS funds and the average fund. We show results for the eight industries with the largest (absolute) differences in industry concentration (relative to the average). The top funds greatly overweight computer and electronic equipment stocks after 2005, and also overweight business services and machinery during shorter spells. Conversely, the top funds underweight retail stocks and, in particular, banking stocks throughout most of the sample.

Very different industry concentrations emerge for the inferior FCS funds (Figure 14). These funds overweight apparel, medical equipment, computers, and wholesale industries during the early 1990s, while conversely banking, business services and pharmaceutical products are underweighted

during this period. Towards the end of the sample, banking, computers, and communication shares get overweighted by the inferior FCS funds.

Overall, this analysis shows that FCS-identified superior funds, while tilting more toward technology-intensive companies, are focused on different sectors of the market during different years. This indicates that the turnover of funds in the top FCS portfolio, over time, that was documented earlier in this paper is at least partially due to the time-varying skills of managers who invest in a particular sector of the market. Moreover, it also suggests that it is difficult for a manager to maintain his performance over long periods of time.

## 7. Conclusion

This paper presents a new approach to selecting funds with superior performance. Our approach goes well beyond earlier studies, in that we predict the *set* of funds most likely to outperform in the future. This goal is very different from predicting whether an individual fund might outperform in the future. By conducting a large set of pair-wise performance comparisons across a large set of mutual funds, the approach iteratively eliminates funds with inferior performance. In line with the finding in recent studies that only a relatively narrow—and declining—set of funds is capable of outperforming on a risk-adjusted basis, our results suggest that it is important to choose a stringent procedure that is capable of eliminating funds with inferior performance.

Several insights emerge from our analysis. First, we find that the set of funds identified *ex ante* to have superior performance subsequently goes on to generate risk-adjusted returns that are substantially higher than the returns obtained by alpha-ranked funds in the top 5-10%, or by funds with significantly positive alphas. Clearly there is substantial heterogeneity in performance even among the funds with the highest alpha estimates.

Second, because of the considerable sampling error surrounding estimates of individual funds' alphas, our results show that it can be beneficial to combine funds' return records with information obtained from their holdings to obtain sufficiently accurate performance estimates making it possible to discriminate superior from inferior funds.

Third, the proportion of funds—as well as the identify of the individual funds—deemed to be superior varies considerably over time and generally is far smaller than the 5-10% of funds in existence often assumed in studies that use portfolio decile sorts to gauge top funds' performance. Consistent with predictions from the theoretical analysis in Kacperczyk et al. (2016), we find that the cross-sectional dispersion in fund performance increases in states with high market volatility and that the performance of funds identified as being superior also rises when aggregate (or idiosyncratic) volatility is high. This suggests that the payoffs to mutual funds from allocating more attention to the stock market is higher during such periods.

Fourth, the performance of the funds identified as top performers, and hence included in the

superior fund confidence set, is notably higher during periods with high aggregate or idiosyncratic volatility than during periods with low volatility, suggesting that the best managers are able to exploit more volatile markets to enhance their risk-adjusted returns.

Fifth, and finally, superior funds achieved their high returns by substantially deviating from the average fund's industry concentration and loadings on systematic risk factors. For example, the superior funds overweighted value stocks from 2003 to 2009, switched between over- and underweighting small cap stocks, but generally took only small bets on momentum risk. These funds also undertook sizeable bets on individual sectors, overweighting computer and electronic equipment stocks but underweighted retail and banking stocks.

## Internet Appendix A: Estimation and prediction of fund performance

This appendix explains how we estimate the unknown parameters of the latent state models and generate predictions of the conditional alpha.

For each fund,  $i$ , we observe a sample of excess returns,  $R_{it}$ . We cast the return model into state space form as follows:

$$\begin{aligned} R_{it} &= a_i F_{it-1} + b_i F_{it-1}^2 - k_i + (\bar{\beta}_i' + \mathbf{c}_i' F_{it-1}) \mathbf{z}_t + \varepsilon_{it} \\ &= G_{it}(F_{it-1}) + \varepsilon_{it}, \\ F_{it} &= \nu_i F_{it-1} + \eta_{it}. \end{aligned}$$

We focus on models where  $R_{it}$  is a linear function of the signal. Define  $\hat{F}_{it|t-1}$  and  $P_{it|t-1}$  as the conditional mean and variance of the  $i$ th fund's signal, given information at time  $t-1$ . The extended Kalman filter relies on a linear approximation of  $G_{it}(F_{it-1})$  around  $\hat{F}_{it-1|t-2}$ ,

$$G_{it}(F_{it-1}) \approx G_{it}(\hat{F}_{it-1|t-2}) + G_{F,it}(F_{it-1} - \hat{F}_{it-1|t-2}),$$

where

$$G_{F,it} = \left. \frac{\partial G_{it}(F_{it-1})}{\partial F_{it-1}} \right|_{F_{it-1} = \hat{F}_{it-1|t-2}}.$$

Given starting values for  $\hat{F}_{i1|0}$  and  $P_{i1|0}$ , the following recursions constitute the extended Kalman filter:

$$\begin{aligned} v_{it} &= R_{it} - G_{it}(\hat{F}_{it-1|t-2}), \\ \hat{F}_{i,t-1|t-1} &= \hat{F}_{it-1|t-2} + P_{it-1|t-2} G'_{F,it-1} K_{it-1}^{-1} v_{it-1}, \\ \hat{F}_{it|t-1} &= \nu_i \hat{F}_{i,t-1|t-1}, \\ K_{it} &= G_{F,it} P_{it-1|t-2} G'_{F,it} + \varepsilon_{it}^2, \\ P_{i,t-1|t-1} &= P_{it-1|t-2} - P_{it-1|t-2} G'_{F,it-1} K_{it-1}^{-1} G_{F,it-1} P_{it-1|t-2}, \\ P_{it|t-1} &= \nu_i^2 P_{i,t-1|t-1} + \eta_{it-1}^2. \end{aligned}$$

Using information up to time  $t-1$ , we can estimate the parameters of the latent state models presented in Section 3. Let  $\hat{\theta}_{i,t-1}$  denote the parameter estimates based on time  $t-1$  information. We use the Kalman filter to forecast the signal one step ahead:

$$\hat{F}_{i,t|t-1} = \hat{F}_{i,t|t-1}(\hat{\theta}_{i,t-1}).$$

For each fund,  $i = 1, \dots, N_t$ , we also predict the alpha one step ahead

$$\hat{\alpha}_{i,t|t-1} = \hat{\alpha}_{i,t|t-1}(\hat{F}_{i,t|t-1}).$$

The forecast of alpha at time  $t$ , given information at time  $t - 1$ , is therefore a function of the forecast of the signal and the parameter estimates available at time  $t - 1$ .

## Internet Appendix B: Local estimates of fund performance

This appendix shows how we generate non-parametric estimates of local (in time) return performance.

To track variation in the performance of portfolios over time, we use a flexible, nonparametric approach that allows for time-varying alpha performance, as well as time-varying factor exposures through the following smooth time-varying parameter model:

$$R_{pt} = \alpha_p(t/T) + \beta_p'(t/T)\mathbf{z}_t + \varepsilon_{pt}, \quad t = 1, \dots, T. \quad (23)$$

Here  $\alpha_p(\cdot)$  and  $\beta_p(\cdot)$  are unknown smooth functions that are allowed to depend on the sample “fraction”,  $t/T$ , and, thus, can vary over time.<sup>48</sup> To see how we can nonparametrically estimate the parameters, define the vector of regressors  $\mathbf{X}_t = (1 \ \mathbf{z}_t)$  and parameters  $\boldsymbol{\theta}_p(t/T) = (\alpha_p(t/T) \ \beta_p'(t/T))'$ , and rewrite equation (23) as

$$R_{pt} = \mathbf{X}_t' \boldsymbol{\theta}_p(t/T) + \varepsilon_{pt}. \quad (24)$$

The parameters  $\boldsymbol{\theta}_p(t/T)$  can be estimated by means of a two-step procedure that first considers the OLS estimator  $\hat{\boldsymbol{\gamma}}_p = (\hat{\boldsymbol{\gamma}}_{p0}, \hat{\boldsymbol{\gamma}}_{p1})'$  of the transformed model

$$k_{st}^{1/2} R_{ps} = k_{st}^{1/2} \mathbf{X}_s' \boldsymbol{\gamma}_{p0} + k_{st}^{1/2} \left( \frac{s-t}{T} \right) \mathbf{X}_s' \boldsymbol{\gamma}_{p1} + \varepsilon_{ps}, \quad s = 1, \dots, T, \quad (25)$$

where  $k_{st} = k(\frac{s-t}{Th})$  is a kernel function, and  $h$  is the bandwidth.<sup>49</sup> In the second step, we construct an estimator of  $\boldsymbol{\alpha}_{pt} = \boldsymbol{\alpha}_p(t/T)$  as

$$\hat{\boldsymbol{\alpha}}_{pt} = (\mathbf{e} \otimes \mathbf{I}) \hat{\boldsymbol{\gamma}}_p, \quad (26)$$

where  $\mathbf{e} = (1, \mathbf{0})$ ,  $\mathbf{I}$  is an identity matrix, and  $\otimes$  denotes the Kronecker product; see Cai (2007) and Chen and Hong (2012) for further discussion of this approach, and its ability to capture time variation in parameter estimates.

<sup>48</sup>Technically, the  $\alpha_p(\cdot)$  and  $\beta_p(\cdot)$  functions allow for a finite number of discontinuities.

<sup>49</sup>Following Chen and Hong (2012), we set  $h = T^{-1/5}/\sqrt{12}$ .



The estimate  $\hat{\alpha}_{p,t}$  from (26) portrays the evolution in the performance of the different portfolios in a way that does not “average out” potentially interesting time-variation in  $\alpha_p$ . This can be contrasted with the conventional full-sample approach or even a rolling-window procedure which does not take into account how much the performance varies over time.

## References

- Avramov, D., Wermers, R., 2006. Investing in mutual funds when returns are predictable. *Journal of Financial Economics* 81, 339–377.
- Banegas, A., Gillen, B., Timmermann, A., Wermers, R., 2013. The cross-section of conditional mutual fund performance in European stock markets. *Journal of Financial Economics* 108, 269–726.
- Barras, L., O. Scaillet, and R. Wermers, 2010, False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas. *Journal of Finance* 65, 179-216.
- Berk, J.B. and R. C. Green, 2004, Mutual fund flows and performance in rational markets. *Journal of Political Economy* 112, 1269-1295.
- Blake, D., A.G. Rossi, A. Timmermann, I. Tonks, and R. Wermers, 2013, Decentralized Investment Management: Evidence from the Pension Fund Industry. *Journal of Finance* 68, 1133-1178.
- Bollen, N., Busse, J., 2004. Short-term persistence in mutual fund performance. *Review of Financial Studies* 18, 569–597.
- Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics* 136, 163–188.
- Carhart, M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 569–597.
- Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* 80, 1157–1183.
- Christopherson, J., Ferson, W., Glassman, D., 1998. Conditioning manager alphas on economic information: Another look at the persistence of performance. *Review of Financial Studies* 11, 111–142.
- Cremers, M. and A. Petajisto, 2009, How Active is Your Fund Manager? A New Measure That Predicts Performance. *Review of Financial Studies* 22(8), 3329-3365.
- Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52, 1035–1058.
- Doornik, J., 2006. *Object-Oriented Matrix Programming Using Ox*, 3rd Edition. London: Timberlake Consultants Press and Oxford: [www.doornik.com](http://www.doornik.com).
- Ferson, W.E., and Schadt, R.W., 1996, Measuring fund strategy and performance in changing economic conditions. *Journal of Finance* 51, 425-461.

- Garleanu, N., and L.H. Pedersen, 2016, Efficiently Inefficient Markets for Assets and Asset Management. Manuscript, UC Berkeley and Copenhagen Business School.
- Glode, V., Hollifield, B., Kacperczyk, M., Kogan, S., 2011. Time-varying predictability in mutual fund returns. Unpublished Manuscript, Wharton School.
- Hansen, P., Lunde, A., 2010. Mulcom 2.0, econometric toolkit for multiple comparisons.
- Hansen, P., Lunde, A., Nason, J., 2011. The model confidence set. *Econometrica* 2 (79), 453–497.
- Harvey, C.R., and Y. Liu, 2016, Rethinking Performance Evaluation. Unpublished manuscript, Duke University.
- Hoberg, G., N. Kumar, and N. Prabhala, 2017. Mutual fund competition, managerial skill, and alpha persistence, *Review of Financial Studies*, forthcoming.
- Kacperczyk, M., C. Sialm, and L. Zheng, 2005, On the Industry Concentration of Actively Managed Equity Mutual Funds. *Journal of Finance* 60, 1983-2012.
- Kacperczyk, M., C. Sialm, and L. Zheng, 2008, Unobserved Actions of Mutual Funds. *Review of Financial Studies* 21, 2379-2416.
- Kacperczyk, M., van Nieuwerburgh, S., Veldkamp, L., 2014. Time-varying fund manager skill. *Journal of Finance* 69, 1455-1484.
- Kacperczyk, M., S van Nieuwerburgh, and L. Veldkamp, 2016, A Rational Theory of Mutual Funds' Attention Allocation. *Econometrica* 84(2), 571-626.
- Kinnel, R., 2010. How expense ratios and star ratings predict success. *Morningstar Advisor* 10, August.
- Kosowski, R., 2011. Do mutual funds perform when it matters most to investors. *Quarterly Journal of Finance* 1, 607–664.
- Liu, J., Timmermann, A., 2013. Optimal convergence trade strategies. *Review of Financial Studies* 26 (4), 1048–1086.
- Mamaysky, H., Spiegel, M., Zhang, H., 2007. Improved forecasting of mutual fund alphas and betas. *Review of Finance* 11, 359–400.
- Mamaysky, H., Spiegel, M., Zhang, H., 2008. Estimating the dynamics of mutual fund alphas and betas. *Review of Financial Studies* 21 (1), 233–264.
- Morey, M., 2005. The kiss of death: A 5-star Morningstar rating?, *Journal of Investment Management* 3 (2).

- Pastor, L., R.F. Stambaugh, and L.A. Taylor, 2015, Scale and Skill in Active Management. *Journal of Financial Economics* 116, 23-45.
- Patton, A., Timmermann, A., 2010. Monotonicity in asset returns: New tests with applications to the term structure, the CAPM and portfolio sorts. *Journal of Financial Economics* 98, 605–625.
- Romano, J., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* 73 (4), 1237–1282.
- Wei, K., R. Wermers, and T. Yao, 2015, Uncommon Value: The Investment Performance of Contrarian Funds. *Management Science*, October, 2394–2414.
- Welch, I., and A. Goyal, 2008. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* 21 (4), 1455–1508.
- Wermers, R., 2000, Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses. *Journal of Finance* 55, 1655-1695.
- White, H., 2000. A reality check for data snooping. *Econometrica* 68 (5), 1097–1126.

**Table 1:** Example of pair-wise comparison

<b>Panel A:</b> alpha (x100)				
	Endowment Inc.	Seligman Com.	Fidelity Tech.	Fidelity Elec.
$\alpha$	0.01	0.02	0.10	0.07
Superior	✓	✓	✓	✓
<b>Panel B:</b> t-statistics of alpha				
	Endowment Inc.	Seligman Com.	Fidelity Tech.	Fidelity Elec.
$t_\alpha$	2.01	1.33	2.32	2.21
Superior	✓	–	✓	✓
<b>Panel C:</b> Pair-wise t-statistics (FCS methodology)				
Step 1 - All				
	Endowment Inc.	Seligman Com.	Fidelity Tech.	Fidelity Elec.
Endowment Inc.	0			
Seligman Com.	0.49	0		
Fidelity Tech.	2.05	1.93	0	
Fidelity Elec.	1.87	2.03	-0.66	0
Step 2 - Endowment Inc. removed				
	Seligman Com.	Fidelity Tech.	Fidelity Elec.	
Seligman Com.	0			
Fidelity Tech.	1.93	0		
Fidelity Elec.	2.03	-0.66	0	
Step 3 - Endowment Inc. and Seligman Com. removed				
	Fidelity Tech.	Fidelity Elec.		
Fidelity Tech.	0			
Fidelity Elec.	-0.66	0		
	Endowment Inc.	Seligman Com.	Fidelity Tech.	Fidelity Elec.
Superior	–	–	✓	✓

The table presents an example of how different fund selection methodologies identify superior funds applied to actual data for four funds obtained in July 2002. For each fund we have a time series of estimated predictive alpha collected over the previous 60 months. The three different panels correspond to different methods of determining which of the funds are superior. In all panels a checkmark, ✓, is used to indicate that a fund has been identified as superior. Panel A presents the average predictive alpha of each fund and labels a fund as superior if the average is positive. In Panel B the selection is based on a standard significance test, where the null hypothesis is that the predictive alpha is zero. Funds are labelled superior if the null is rejected at a 5% significance level. Panel C is based on a series of pair-wise tests and divided into three steps. In each step the fund in the row is tested against the fund in the column and the column fund, which is dominated most by a row fund (i.e., generates the highest t-statistic) is eliminated. The methodology labels funds as superior, which cannot be eliminated based on a pre-determined level of significance.

**Table 2:** Cross-section of alpha estimates

	Latent Skill Holdings
Mean	-0.066
Min	-1.610
Max	1.713
5%	-0.376
10%	-0.273
25%	-0.157
50%	-0.062
75%	0.034
90%	0.136
95%	0.211

This table shows the cross-sectional distribution of four-factor alpha estimates (in percent per month) obtained from monthly returns data on U.S. equity mutual funds over the period 1980:06-2012:12. Mutual fund returns are net of transaction costs and fees and are calculated in excess of a one-month T-bill rate. The four risk factors are excess returns on a market portfolio, small-minus-big market cap and high-minus-low book-to-market value Fama-French factors and a momentum risk factor. The table reports results for a latent skill-holdings (LSH) model, which allows for a time-varying alpha and combines returns and holdings data to estimate each fund's alpha. The extended Kalman filter is used to extract the latent signal underlying the model. The reported estimates of the cross-sectional distribution is based on the individual funds' average alpha estimates.

**Table 3:** Risk-adjusted performance for alpha-ranked decile portfolios

Portfolio	Latent State Holdings	
P1A	-0.18	(-2.64)
P1B	-0.14	(-2.39)
P1	-0.16	(-2.66)
P2	-0.14	(-2.66)
P3	-0.10	(-2.20)
P4	-0.09	(-2.03)
P5	-0.09	(-2.37)
P6	-0.06	(-1.52)
P7	-0.05	(-1.37)
P8	-0.02	(-0.46)
P9	0.00	(0.05)
P10	0.02	(0.35)
P10A	-0.00	(-0.04)
P10B	0.05	(0.65)
P10-P1	0.18	(4.01)
MR	0.00	

This table reports four-factor alphas (in percent per month) for a set of alpha-ranked decile portfolios, with t-statistics in brackets. Each month we rank the set of mutual funds in existence according to their expected alphas and allocate them to equal-weighted decile portfolios listed in increasing order from funds with the lowest alphas ( $P1$ ) to funds with the highest alphas ( $P10$ ). The alpha ranking is repeated each month during the sample 1990:07-2012:12 and so produces a time series of portfolio returns from which the reported alpha estimates are computed. The table reports results the latent skill-holdings model, which allows for a time-varying alpha and combines returns and holdings data to estimate each fund's alpha.  $P1A$  and  $P1B$  are constructed by splitting the funds in the bottom decile portfolio into two new portfolios of equal size.  $P10A$  and  $P10B$  are constructed in the same way.  $P10 - P1$  is a portfolio constructed from the difference between the top and bottom decile portfolios. The table reports alpha estimates in percentage terms followed by t-statistics in brackets. If a particular performance model provides an accurate ranking of the funds we would expect the alpha estimates to increase monotonically from  $P1$  to  $P10$ . We test for this monotonically increasing pattern against a flat or a decreasing relation using the monotonic relation (MR) test proposed by Patton and Timmermann (2010). Low p-values are evidence against a flat or decreasing pattern and so are suggestive of a precise ranking of alphas.

**Table 4:** Risk-adjusted performance for FCS portfolios of superior funds

	Latent State Holdings	
	Superior funds	Inferior funds
FCS-tight	0.48 (2.36)	-0.23 (-2.33)
FCS-medium-tight	0.37 (2.14)	-0.18 (-2.24)
FCS-medium	0.14 (1.10)	-0.15 (-2.08)
FCS-wide	0.13 (1.10)	-0.20 (-3.67)
P10/P1	0.02 (0.35)	-0.16 (-2.67)

This table reports four-factor alphas for a set of portfolios formed by equal-weighting funds identified by the fund confidence set (FCS) approach. Alpha estimates are in percent per month with t-statistics shown in brackets. Each month we apply the FCS approach to all funds determined to have significantly positive (negative) predictive alpha by the Romano Wolf first stage filter as to identify the set of superior (inferior) funds. We then form a portfolio of these funds and record its return during the following month. This procedure is repeated each month during the sample 1985:07-2012:12 and so produces a time series of portfolio returns from which the reported four-factor alpha estimates are computed. The top four rows report results for four different values of the tightness parameter ( $\lambda$ ) used to determine how strict to be when eliminating funds from the FCS. The top row uses a tight choice ( $\lambda = 0.90$ ), resulting in a narrower set of funds being included, while the second, third, and fourth rows give rise to medium-tight ( $\lambda = 0.75$ ), medium ( $\lambda = 0.50$ ) and wide ( $\lambda = 0.10$ ) fund confidence sets. For comparison we also show results for the alpha-ranked top decile portfolio ( $P10$ ) for superior funds and the alpha-ranked bottom portfolio ( $P1$ ) for inferior funds. The columns present results for both superior and inferior funds based on the latent skill-holdings (LSH) model, which allows for a time-varying alpha and combines returns and holdings data to estimate each fund's alpha.



**Table 5:** Dispersion and risk-adjusted performance sorted by volatility

<b>Panel A:</b> Dispersion in Industry Concentration and CS Measure				
	Industry concentration		CS measure	
	Aggregate risk		Aggregate risk	
	All funds		All funds	
Low risk	0.464 [0.000]		2.291 [0.000]	
Med. risk	0.680 [0.000]		3.399 [0.000]	
High risk	1.023 [0.000]		6.093 [0.000]	
High-Low	0.559 [0.023]		3.802 [0.000]	

<b>Panel B1:</b> Dispersion of performance for superior funds (12 months)				
	Aggregate risk		Idiosyncratic risk	
	FCS	P10	FCS	P10
	Low risk	-0.009 [0.896]	0.005 [0.834]	-0.023 [0.783]
Med. risk	0.217 [0.024]	-0.094 [0.004]	0.082 [0.334]	-0.060 [0.041]
High risk	0.639 [0.000]	-0.016 [0.700]	0.788 [0.000]	0.019 [0.671]
High-Low	0.649 [0.000]	-0.021 [0.661]	0.811 [0.000]	0.083 [0.104]

<b>Panel B2:</b> Dispersion of performance for superior funds (36 months)				
	Aggregate risk		Idiosyncratic risk	
	FCS	P10	FCS	P10
	Low risk	0.117 [0.000]	-0.017 [0.169]	0.071 [0.016]
Med. risk	0.472 [0.000]	-0.017 [0.394]	0.219 [0.000]	-0.107 [0.000]
High risk	0.816 [0.000]	0.061 [0.006]	1.115 [0.000]	0.124 [0.000]
High-Low	0.699 [0.000]	0.078 [0.002]	1.043 [0.000]	0.114 [0.000]

Panel A of this table reports dispersion in industry concentrations and CS measure across all funds sorted into groups according to volatility. We define the aggregate risk in period  $t$  as  $|\beta_t \sigma_{mt}|$ , where  $\beta_t$  is the average market beta across funds and  $\sigma_{mt}$  is the realized volatility of the market based on daily returns. The idiosyncratic risk is calculated in a similar fashion. The risk measures are based on a rolling window with a length of 12 monthly observations. Dispersion in industry concentration is calculated as  $\sum_{c=1}^{48} (w_{i,t}^c - w_{avg,t}^c)^2$ , where  $w_{i,t}^c$  is fund  $i$ 's weight in industry  $c$  at time  $t$ , and  $w_{avg,t}^c$  is the weight of the average fund in industry  $c$  at time  $t$ . The dispersion in the CS measure is calculated as the average squared deviation from a fund's CS measure to the average CS measure across funds. Panel B1 reports average alpha performance of equal-weighted portfolios of superior funds identified by the fund confidence set (FCS) and the decile approach (P10) sorted according to the rolling 12-month volatility estimate. Average alpha estimates are reported in percent per month. Each month we apply the FCS approach to all funds with significantly positive predictive alpha as determined by the Romano Wolf first stage filter in order to identify the set of superior funds. We then form a portfolio of these funds and record its return during the following month. This procedure is repeated each month during the sample 1985:07-2012:12 and so produces a time series of portfolio returns from which four-factor alpha estimates are computed based on the same 12-month rolling window used to calculate the volatility measures. Panel B2 repeats this analysis for a 36-month rolling window.

**Table 6:** Width of fund confidence set and macroeconomic state variables

State variable	Superior Funds		Inferior Funds	
	Percentage of funds		Percentage of funds	
Constant	-0.00 (-0.35)		0.12 (2.44)	
tbl	0.45 (3.77)		0.76 (2.91)	
tms	1.23 (6.01)		0.30 (1.02)	
dfy	-0.88 (-3.17)		1.33 (1.87)	
dy	0.00 (0.89)		0.04 (3.15)	

This table presents least squares estimates (with t-statistics in parentheses) from regressing the percentage of funds identified by the tight FCS approach as being superior (left panel) or inferior (right panel). The FCS uses as the initial set of candidate funds those funds identified to have significantly positive alpha estimates by the Romano-Wolf (2005) approach and assumes that  $\lambda = 0.90$ . We use the latent skill-holdings model with a time-varying alpha to compute each fund's alpha estimate and thus use both returns and holdings data. The macroeconomic state variables are the one-month T-bill rate (tbl), the term spread (tms), the default yield spread (dfy), and the dividend yield (dy).

**Table 7:** Attributes of funds included in the Fund Confidence Set

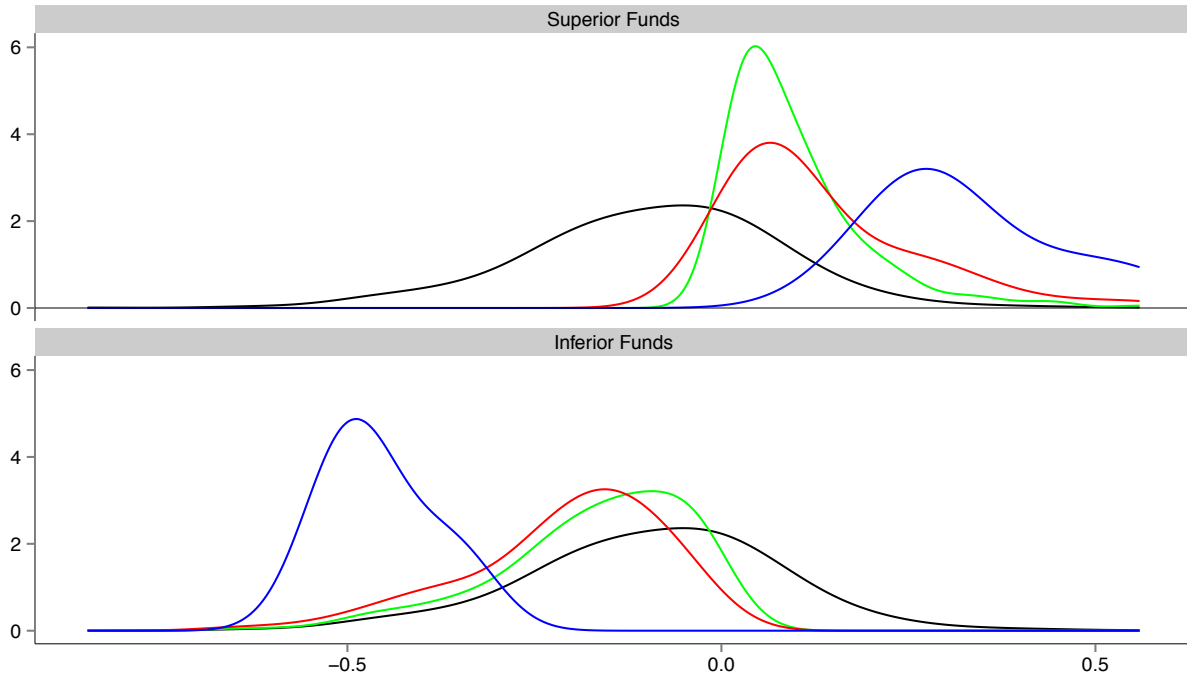
	Overall			High vol.			Low vol.		
	Superior	Inferior	Average	Superior	Inferior	Average	Superior	Inferior	Average
CS-1m( $\times 100$ )	0.33	-0.04	0.03	0.30	0.00	0.02	0.24	0.08	0.08
CS-3m( $\times 100$ )	0.38	-0.12	0.04	0.51	-0.11	0.04	0.24	-0.09	0.05
TNA	3884.23	636.23	1765.90	3105.27	1037.47	1997.49	4301.62	346.97	1514.06
Size	4.18	4.01	4.14	4.35	4.07	4.20	4.08	3.76	4.07
BTM	2.80	2.95	2.89	2.67	2.90	2.86	2.89	2.99	2.91
MOM	3.23	3.19	3.14	3.15	3.18	3.10	3.28	3.13	3.13
RetGap12( $\times 100$ )	0.04	-0.04	-0.04	0.17	0.03	0.04	-0.05	-0.17	-0.08
Flow( $\times 100$ )	2.90	-0.63	1.40	2.12	-2.43	-0.45	2.85	0.41	2.57
ExpRatio( $\times 100$ )	1.11	1.58	1.22	1.03	1.43	1.20	1.16	1.76	1.25
Turnover	1.05	1.35	0.83	1.32	1.30	0.86	0.75	1.19	0.80

This table reports several attributes of the funds selected by the FCS approach. The table reports attributes for the FCS portfolio of superior funds, the FCS portfolio of inferior funds and the average across funds. The FCS portfolios are based on a tight parameterization ( $\lambda = 0.9$ ) and the funds selected by the Romano Wolf first stage filter, using the latent skill holdings model to estimate time-varying alphas. Attributes are calculated by first taking cross-sectional averages of attributes of the funds in the portfolio at a given point in time, then taking time series averages. The attributes include the portfolio-weighted characteristic selectivity measure, one month (CS-1m) and three months (CS-3m) in percent, total net assets of funds in millions of dollars (TNA), portfolio holdings-based style attributes in the size (Size), book-to-market (BTM), and momentum (MOM) dimensions, lagged net return, compounded over the 12 months prior to each portfolio formation date (RetGap12), monthly percentage net cash inflows (Flow), computed as TNA minus one-quarter-lagged TNA, divided by three, fund expense ratio (ExpRatio), and percent monthly turnover (Turnover). These attributes are presented for all periods, for periods with high aggregate volatility only and for periods with low aggregate volatility only. Aggregate volatility is calculated as in Table 5.

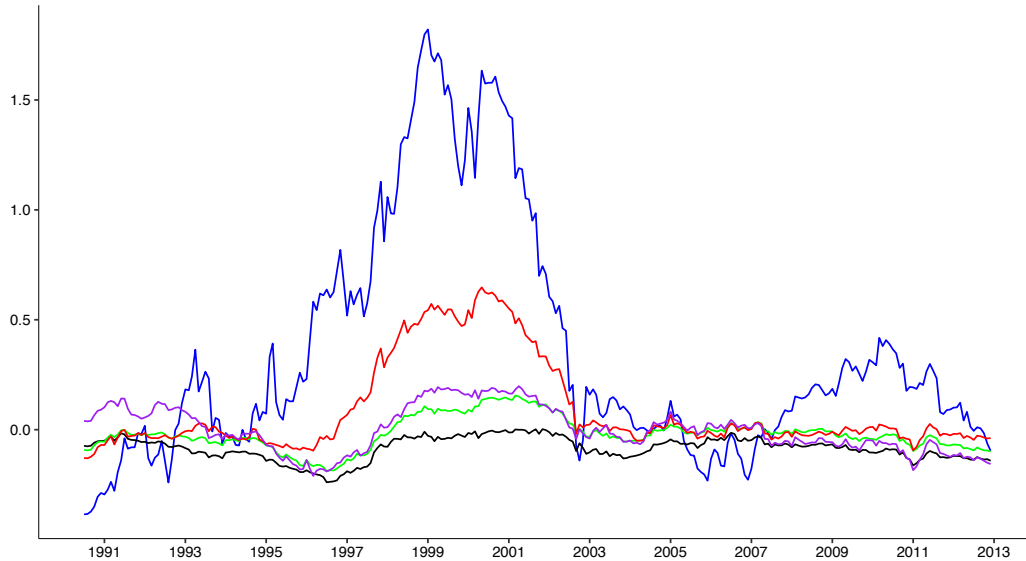
**Table 8:** Industry concentrations of funds included in the fund confidence set

	Overall			High vol.			Low vol.		
	Superior	Inferior	Average	Superior	Inferior	Average	Superior	Inferior	Average
Agriculture	0.15	0.18	0.21	0.07	0.28	0.26	0.09	0.12	0.18
Food Products	0.93	1.43	1.56	0.46	1.38	1.38	1.10	1.35	1.63
Candy & Soda	0.16	0.26	0.11	0.25	0.26	0.13	0.13	0.18	0.09
Beer & Liquor	0.68	1.07	0.95	0.38	0.94	0.96	1.20	1.30	0.96
Tobacco Products	0.26	0.25	0.34	0.32	0.26	0.38	0.23	0.23	0.31
Recreation	0.22	0.34	0.34	0.14	0.24	0.25	0.27	0.48	0.41
Entertainment	1.75	0.78	0.99	2.15	0.63	0.72	1.77	0.87	1.21
Printing & Publishing	1.42	0.87	1.00	1.34	0.71	0.79	1.22	0.95	1.14
Consumer Goods	1.41	2.89	1.98	1.05	2.61	1.87	1.96	3.10	2.07
Apparel	0.48	1.81	0.85	0.27	0.98	0.81	0.71	3.18	0.91
Healthcare	0.68	1.44	1.31	0.29	1.06	1.10	1.06	1.72	1.45
Medical Equipment	1.30	2.39	2.13	0.64	1.88	2.14	1.39	3.00	2.10
Pharmaceutical Products	4.36	6.86	7.01	2.46	7.51	7.63	5.53	5.90	6.24
Chemicals	1.15	2.04	2.36	0.55	1.38	1.91	1.82	2.54	2.73
Rubber and Plastic Products	0.22	0.27	0.33	0.12	0.28	0.27	0.18	0.30	0.37
Textiles	0.14	0.23	0.25	0.06	0.08	0.14	0.24	0.33	0.36
Construction Materials	0.64	0.86	0.97	0.63	0.66	0.83	0.71	1.03	1.10
Construction	0.28	0.95	0.74	0.33	0.75	0.69	0.32	1.28	0.83
Steel Works Etc	0.61	1.38	1.15	0.45	0.97	0.90	0.69	1.68	1.30
Fabricated Products	0.05	0.13	0.10	0.05	0.11	0.07	0.07	0.15	0.13
Machinery	3.16	3.51	2.89	1.67	2.71	2.92	1.98	4.11	2.94
Electrical Equipment	1.09	1.55	1.50	0.84	1.85	1.42	1.64	1.43	1.52
Automobiles and Trucks	1.02	1.42	1.32	0.82	1.24	1.04	1.16	1.66	1.58
Aircraft	0.49	1.30	1.13	0.35	1.17	1.17	0.71	1.49	1.11
Shipbuilding and Railroad Equipment	0.04	0.10	0.17	0.05	0.14	0.21	0.05	0.09	0.16
Defense	0.13	0.29	0.28	0.06	0.34	0.28	0.19	0.32	0.29
Precious Metals	0.23	0.10	1.15	0.17	0.09	0.82	0.27	0.09	1.41
Non-Metallic and Industrial Metal Mining	0.30	0.31	0.65	0.20	0.31	0.64	0.36	0.38	0.71
Coal	0.13	0.12	0.15	0.07	0.10	0.20	0.29	0.19	0.16
Petroleum and Natural Gas	2.95	5.21	5.81	1.94	5.06	6.04	4.27	5.10	5.64
Utilities	1.12	2.96	4.12	0.98	2.54	3.98	1.39	2.79	4.13
Communication	12.20	4.89	4.79	18.53	4.33	5.08	9.39	4.67	4.40
Personal Services	0.66	0.74	0.65	0.42	0.75	0.66	1.01	0.58	0.64
Business Services	13.77	8.27	8.34	18.55	10.22	9.79	10.46	7.05	7.42
Computers	13.24	6.63	4.82	11.06	6.49	5.11	13.87	6.52	4.43
Electronic Equipment	10.60	6.13	5.35	12.29	6.50	5.91	7.65	6.36	5.01
Measuring and Control Equipment	1.44	1.28	1.30	1.04	1.47	1.41	1.16	1.20	1.24
Business Supplies	0.84	1.61	1.61	0.58	1.26	1.33	1.06	1.77	1.73
Shipping Containers	0.11	0.29	0.27	0.07	0.26	0.28	0.15	0.35	0.26
Transportation	1.71	1.83	2.20	1.74	1.68	1.97	1.81	2.10	2.52
Wholesale	1.37	2.36	1.91	0.93	2.58	2.00	1.77	2.52	1.85
Retail	6.03	6.38	6.47	7.13	7.33	6.52	6.14	5.91	6.46
Restaurants and Hotels and Motels	1.08	1.32	1.22	1.17	1.15	1.27	1.15	1.84	1.25
Banking	3.98	7.65	8.47	3.14	9.45	8.29	4.90	5.19	8.60
Insurance	3.45	4.48	5.43	2.32	4.71	5.03	4.26	4.13	5.73
Real Estate	0.15	0.09	0.25	0.25	0.11	0.28	0.08	0.11	0.23
Trading	1.22	2.07	2.37	0.94	2.57	2.46	1.53	1.61	2.41
Almost Nothing	0.60	0.67	0.72	0.67	0.59	0.67	0.58	0.76	0.64

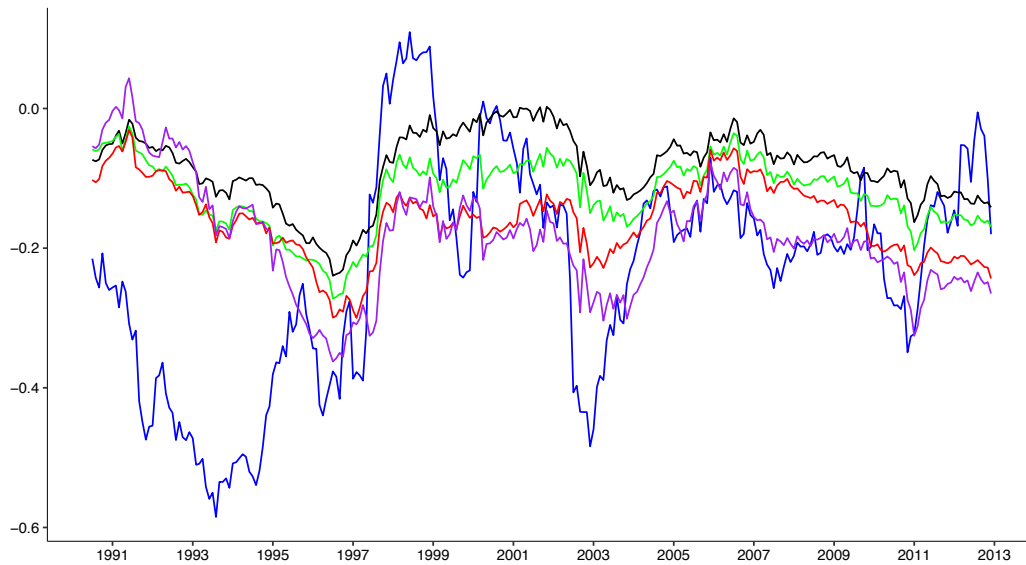
This table reports industry concentrations in percent for the FCS portfolio of superior funds, the FCS portfolio of inferior funds, and the average across funds. The FCS portfolios are based on a tight parameterization ( $\lambda = 0.9$ ) and the funds selected by the Romano Wolf first stage filter, using the latent skill holdings model to estimate time-varying alphas. Average industry concentrations are calculated by first taking cross-sectional averages of the industry concentrations of the funds in the portfolio at a given point in time, then taking time series averages. These industry concentrations are presented for all periods, for periods with high aggregate volatility and for periods with low aggregate volatility only. Aggregate volatility is calculated as in Table 5.



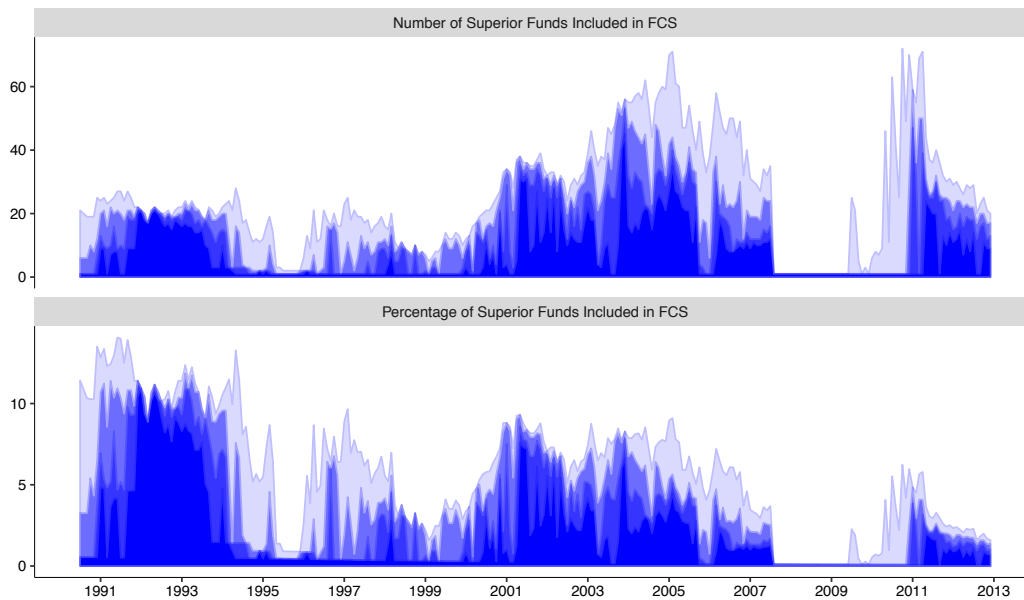
**Figure 1:** Time variation in alpha forecasts for top-ranked funds. The figure presents distributions of predictive alphas for funds included in various portfolios at a single month in our sample (July 2006). The top panel presents distributions of alphas for the top-ranked funds, while the bottom panel shows results for the bottom-ranked funds. In both cases the black line represents a portfolio consisting of all funds in existence in July 2006. The green line shows the distribution for funds with positive (top panel) or negative (bottom panel) predictive alphas. The red line shows the distribution for funds with significantly positive (top panel) or negative (bottom panel) predictive alphas, identified using the Romano Wolf (2005) approach. The blue line shows the distribution of funds included in the superior (top) or inferior (bottom) FCS, in both cases assuming a narrow set ( $\lambda = 0.90$ ) and using the time-varying latent skill and holdings performance model.



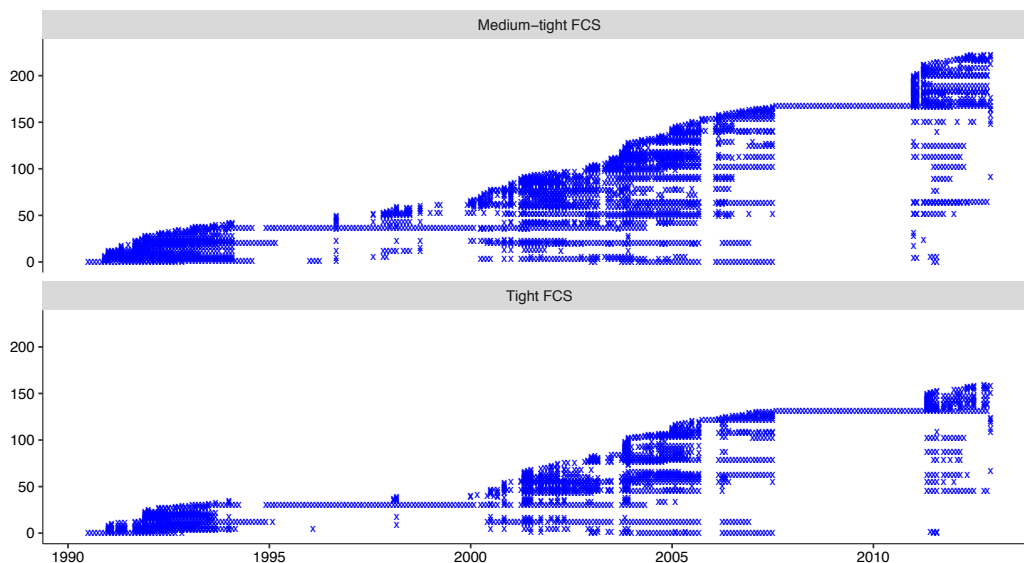
**Figure 2:** Time variation in alpha estimates for superior funds. This figure shows nonparametric estimates of local time-variation in four-factor alpha estimates (in percent per month) for a variety of portfolios consisting of top-ranked funds. The blue line tracks the performance of the tight FCS portfolio ( $\lambda = 0.90$ ) of superior funds based on the latent skill holdings model. The green line tracks the performance of a portfolio consisting of all funds with positive predictive alphas while the red line tracks the performance of a portfolio consisting of the funds with significantly positive alpha estimates. The purple line tracks the performance of a portfolio consisting of the top decile of alpha-ranked funds ( $P10$ ). For comparison the black line tracks the variation in the alpha of the average fund.



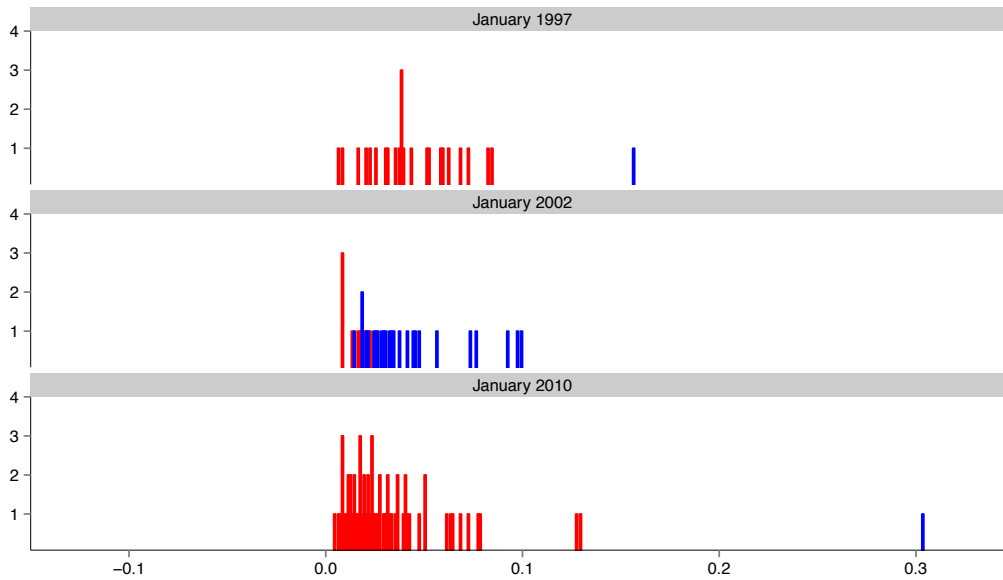
**Figure 3:** Time variation in alpha estimates for inferior funds. This figure shows nonparametric estimates of local time-variation in four-factor alpha estimates (in percent per month) for a variety of portfolios consisting of bottom-ranked funds. The blue line tracks the performance of the tight FCS portfolio ( $\lambda = 0.90$ ) of inferior funds based on the latent skill holdings model. The green line tracks the performance of a portfolio consisting of all funds with negative predictive alphas while the red line tracks the performance of a portfolio consisting of the funds with significantly negative alpha estimates. The purple line tracks the performance of a portfolio consisting of the bottom decile of alpha-ranked funds ( $P1$ ). For comparison the black line tracks the variation in the alpha of the average fund.



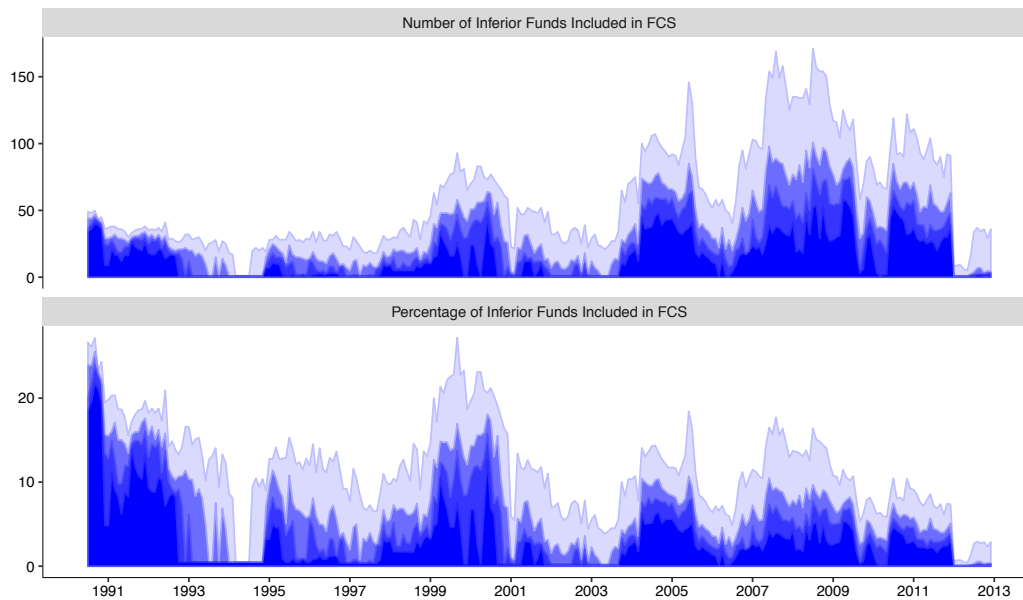
**Figure 4:** Evolution in the set of funds identified as being superior by the FCS approach. The top panel shows the evolution in the number of funds included in the tight, medium and wide fund confidence sets of superior funds. The bottom panel shows the corresponding evolution in the percentage of funds identified as being superior by the FCS approach. Funds are selected from the set of funds with significantly positive predictive alpha using the latent skill-holdings model with a time-varying alpha to compute each fund's alpha estimate. Dark blue areas correspond to  $\lambda = 0.90$  (tight set), while lighter areas correspond to  $\lambda = 0.75$  (medium-tight set),  $\lambda = 0.50$  (medium set), and  $\lambda = 0.10$  (wide set).



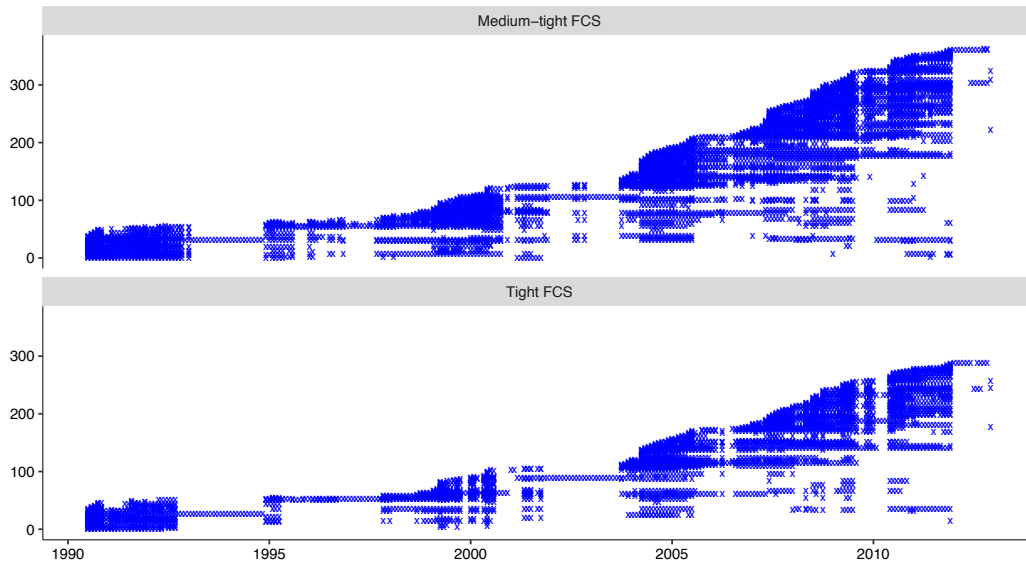
**Figure 5:** Identification of individual funds with superior skill by the FCS approach. The plot illustrates the evolution in the composition of individual funds with superior skills as identified by the FCS approach. Each fund that is included in the FCS at least once during the sample is assigned a unique number on the y-axis based on the date of the first inclusion and a cross shows when this fund is included. The FCS is based on the latent skill holdings model with time-varying alpha, considers funds with significantly positive predictive alpha as candidate funds and assumes  $\lambda = 0.75$  in the top panel and  $\lambda = 0.90$  in the bottom panel.



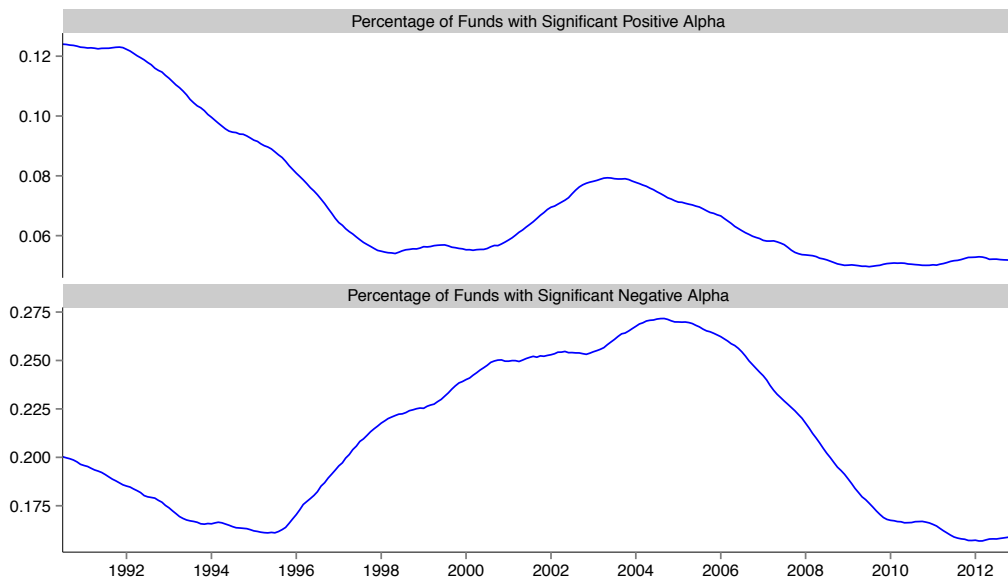
**Figure 6:** Histograms of the average predictive alpha measure for three different periods in the sample. Red bars indicate funds (chosen from the set of funds with significantly positive alpha) which are not included in the FCS assuming that  $\lambda = 0.90$ . Blue bars indicate funds which are included in the FCS.



**Figure 7:** Evolution in the set of funds identified as being inferior by the FCS approach. The top panel shows the evolution in the number of funds included in the tight, medium and wide fund confidence sets of inferior funds. The bottom panel shows the corresponding evolution in the percentage of funds identified as being inferior by the FCS approach. Funds are selected from the set of funds with significantly negative alpha using the latent skill-holdings model with a time-varying alpha to compute each fund's alpha estimate. Dark blue areas correspond to  $\lambda = 0.90$  (tight set), while lighter areas correspond to  $\lambda = 0.75$  (medium-tight set),  $\lambda = 0.50$  (medium set), and  $\lambda = 0.10$  (wide set).

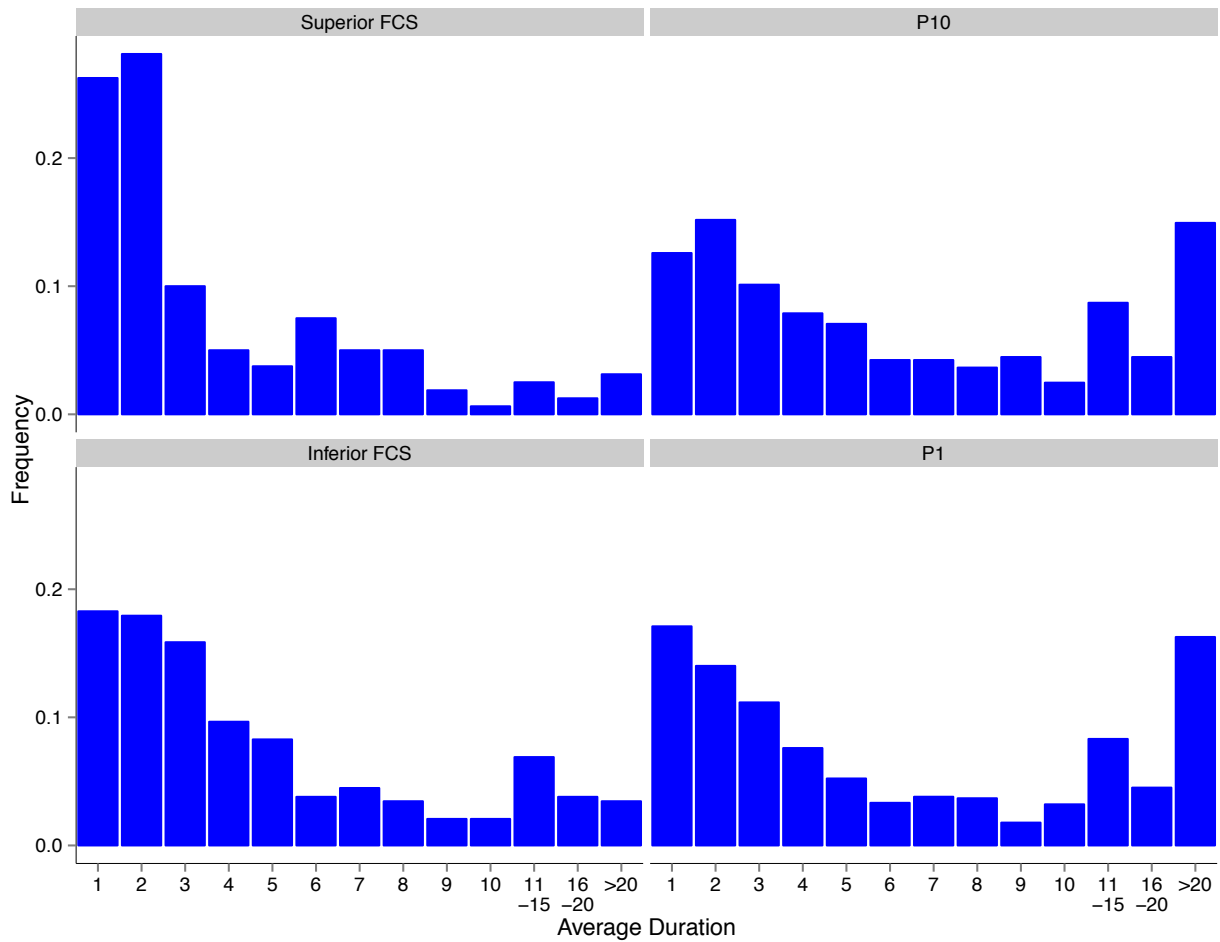


**Figure 8:** Identification of individual funds with inferior skill by the FCS approach. The plot illustrates the evolution in the composition of individual funds with inferior skills as identified by the FCS approach. Each fund that is included in the FCS at least once during the sample is assigned a unique number on the y-axis based on the date of the first inclusion and a cross shows when this fund is included. The FCS is based on the latent skill holdings model with time-varying alpha, considers funds with significantly negative alpha as candidate funds and assumes  $\lambda = 0.75$  in the top panel and  $\lambda = 0.90$  in the bottom panel.

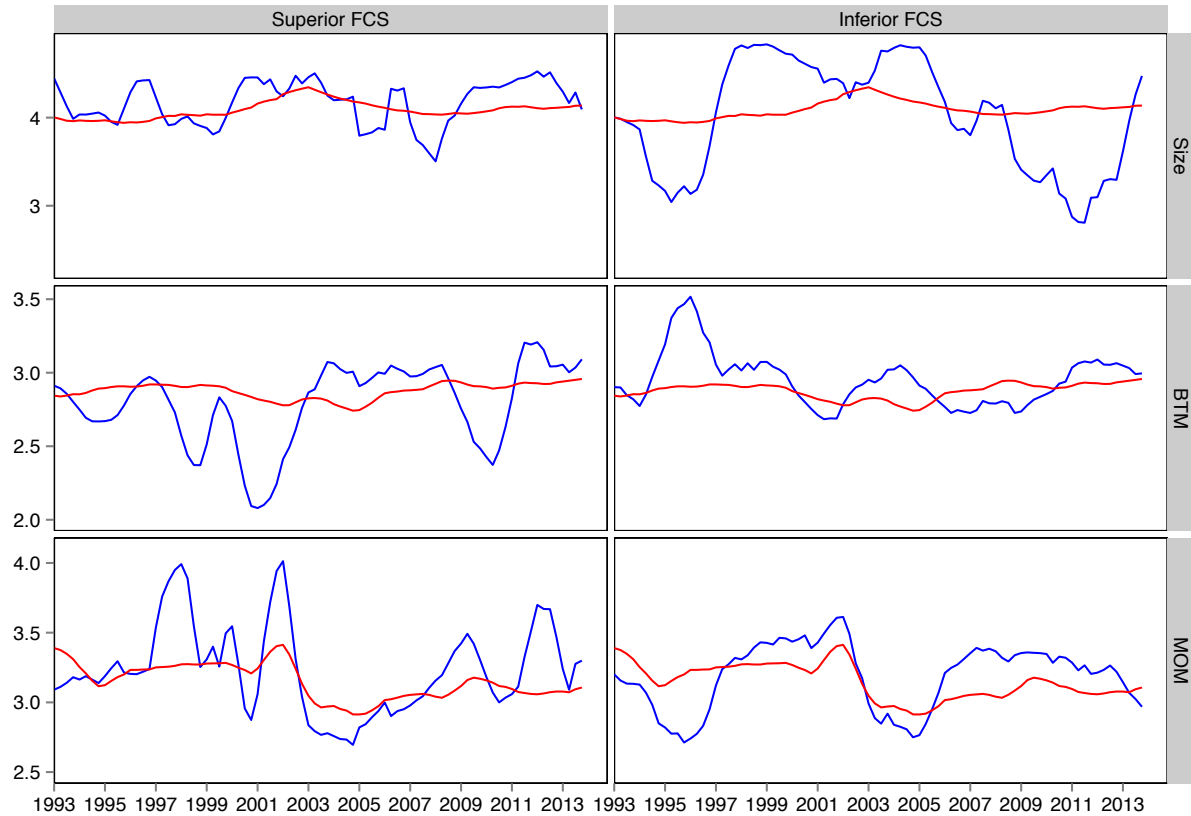


**Figure 9:** Kernel estimate of the evolution in the proportion of funds with significantly positive (top panel) or negative (bottom panel) alpha estimates, using the Romano-Wolf (2005) methodology. Each month we determine the set of funds with significantly positive (or negative) alphas by applying the Romano-Wolf (2005) step-wise bootstrap procedure to the alpha estimates obtained from the latent skill model that combines data on fund returns and quarterly holdings.

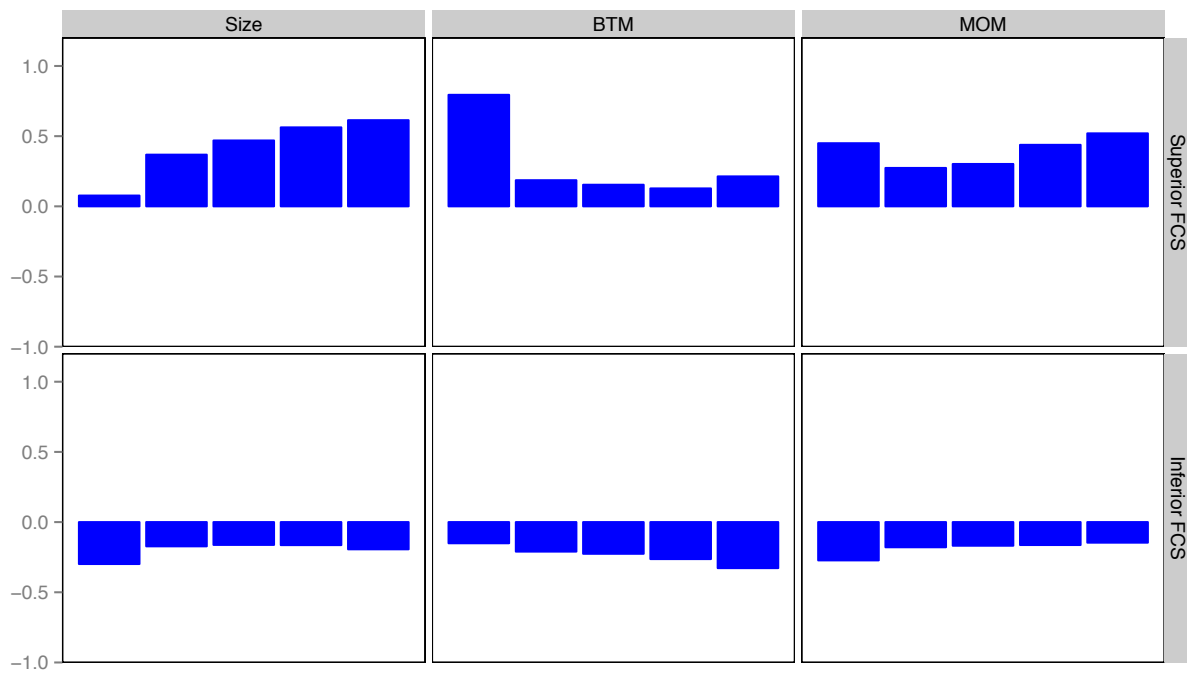




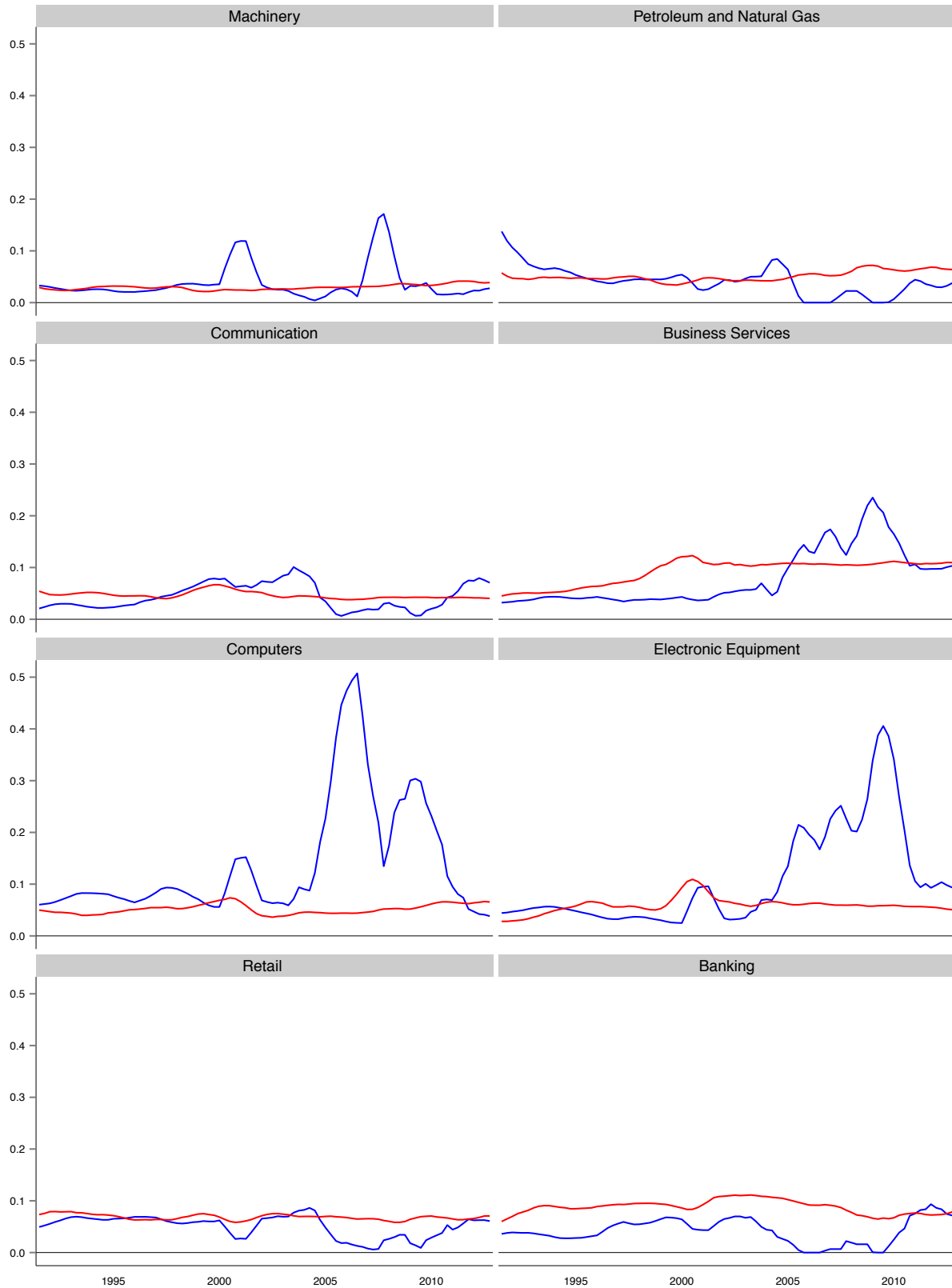
**Figure 10:** Average duration for funds getting included in different portfolios. The histograms show the distribution of the number of consecutive months where individual funds get included in four different portfolios, namely the FCS portfolio of superior (top left corner) and inferior funds (bottom left), the top (top right) and bottom decile (bottom right) portfolios. Only funds that get included in the portfolio at least once are represented in the plot.



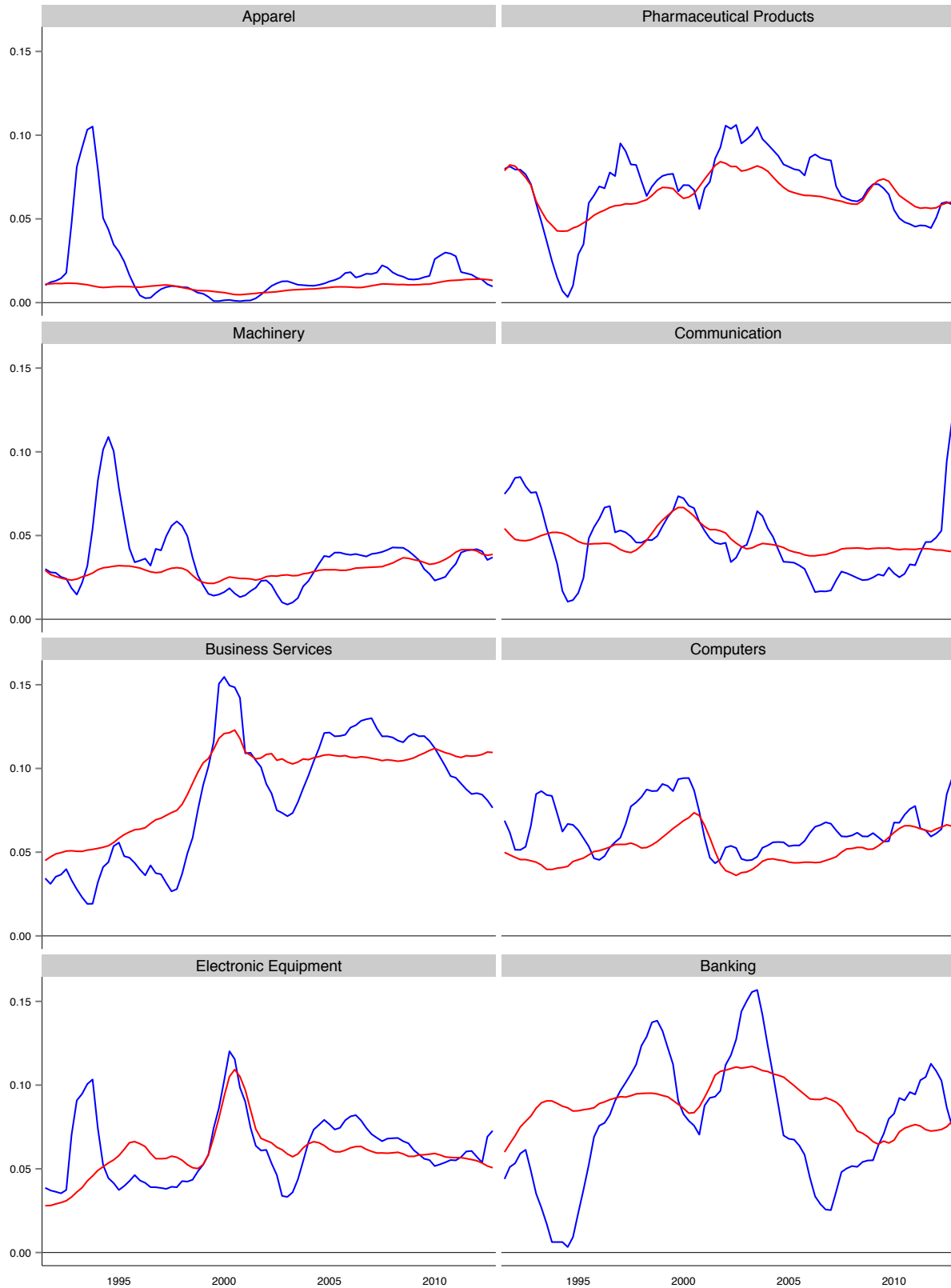
**Figure 11:** Evolution in loadings on risk factors. Blue lines show equal-weighted loadings for funds in the FCS portfolios. Red lines show loadings for the average fund for comparison. The left column shows one-year rolling averages of exposures to the size, book-to-market and momentum risk factors for the FCS portfolio consisting of superior funds. The right column shows one-year rolling averages of exposures to the same risk factors for the FCS portfolio consisting of inferior funds. The plots assume  $\lambda = 0.90$  and are based on the time-varying alpha model that combines data on fund returns and holdings to estimate alphas.



**Figure 12:** Performance of the superior and inferior FCS portfolios for different realizations of the risk factors. The top row shows histograms of the four-factor alphas for the FCS portfolio of superior funds for different values of the size factor (left panel), book-to-market factor (middle panel) and momentum factor (right panel). The bottom row presents four-factor alpha estimates for the FCS portfolio of inferior funds. The plots assume  $\lambda = 0.90$  and are based on the time-varying alpha model that combines data on fund returns and holdings to estimate alphas.



**Figure 13:** Industry concentrations for the FCS portfolio of superior funds. The figure presents one-year rolling averages of industry concentrations for the FCS portfolio composed of funds identified to have superior performance (blue line). We have chosen the eight industries with the highest maximum industry concentration. For comparison the red lines show the average concentration in the same industries computed across all funds in our sample. The plots assume  $\lambda = 0.90$  and are based on the time-varying alpha model that combines data on fund returns and holdings to estimate alphas.



**Figure 14:** Industry concentrations for the FCS portfolio of inferior funds. The figure presents one-year rolling averages of industry concentrations for the FCS portfolio composed of funds identified to have inferior performance (blue line). We have chosen the 8 industries with the highest maximum industry concentration. For comparison the red lines show the average concentration in the same industries computed across all funds in our sample. The plots assume  $\lambda = 0.90$  and are based on the time-varying alpha model that combines data on fund returns and holdings to estimate alphas.

# Research Papers 2016



- 2016-29: Hossein Asgharian, Charlotte Christiansen, Rangan Gupta and Ai Jun Hou: Effects of Economic Policy Uncertainty Shocks on the Long-Run US-UK Stock Market Correlation
- 2016-30: Morten Ørregaard Nielsen and Sergei S. Shibaev: Forecasting daily political opinion polls using the fractionally cointegrated VAR model
- 2016-31: Carlos Vladimir Rodríguez-Caballero: Panel Data with Cross-Sectional Dependence Characterized by a Multi-Level Factor Structure
- 2016-32: Lasse Bork, Stig V. Møller and Thomas Q. Pedersen: A New Index of Housing Sentiment
- 2016-33: Joachim Lebovits and Mark Podolskij: Estimation of the global regularity of a multifractional Brownian motion
- 2017-01: Nektarios Aslanidis, Charlotte Christiansen and Andrea Cipollini: Predicting Bond Betas using Macro-Finance Variables
- 2017-02: Giuseppe Cavaliere, Morten Ørregaard Nielsen and Robert Taylor: Quasi-Maximum Likelihood Estimation and Bootstrap Inference in Fractional Time Series Models with Heteroskedasticity of Unknown Form
- 2017-03: Peter Exterkate and Oskar Knapik: A regime-switching stochastic volatility model for forecasting electricity prices
- 2017-04: Timo Teräsvirta: Sir Clive Granger's contributions to nonlinear time series and econometrics
- 2017-05: Matthew T. Holt and Timo Teräsvirta: Global Hemispheric Temperatures and Co-Shifting: A Vector Shifting-Mean Autoregressive Analysis
- 2017-06: Tobias Basse, Robinson Kruse and Christoph Wegener: The Walking Debt Crisis
- 2017-07: Oskar Knapik: Modeling and forecasting electricity price jumps in the Nord Pool power market
- 2017-08: Malene Kallestrup-Lamb and Carsten P.T. Rosenskjold: Insight into the Female Longevity Puzzle: Using Register Data to Analyse Mortality and Cause of Death Behaviour Across Socio-economic Groups
- 2017-09: Thomas Quistgaard Pedersen and Erik Christian Montes Schütte: Testing for Explosive Bubbles in the Presence of Autocorrelated Innovations
- 2017-10: Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Dynamics of Variance Risk Premia, Investors' Sentiment and Return Predictability
- 2017-11: Søren Johansen and Morten Nyboe Tabor: Cointegration between trends and their estimators in state space models and CVAR models
- 2017-12: Lukasz Gatarek and Søren Johansen: The role of cointegration for optimal hedging with heteroscedastic error term
- 2017-13: Niels S. Grønborg, Asger Lunde, Allan Timmermann and Russ Wermers: Picking Funds with Confidence