Picking Funds with Confidence[*]

Niels S. Grønborg *Aarhus University, CREATES and DFI*
Asger Lunde *Aarhus University and CREATES*
Allan Timmermann *UCSD*
Russ Wermers *University of Maryland*

Abstract

We present a new approach to selecting actively-managed 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, and locates funds with substantially higher risk-adjusted returns than those identified by conventional alpha ranking methods. We find strong evidence of time-series variation in both the number of funds identified as superior using our approach, as well as in their performance across different economic states.

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

JEL codes: G2, G11, G17

1. Introduction

Each year, large sums of money and vast amounts of effort are dedicated towards selecting the best mutual funds.[1] Yet, despite decades of academic research, forecasting which mutual funds will go on to produce the best performance remains an elusive goal. While many studies have singled out different characteristics of individual funds that are correlated with future performance, it is unclear how to effectively identify the *set* of funds that is most likely to outperform in the future.[2] Indeed, the identification of a superior set of funds, rather than an individual fund, involves a considerably more complicated statistical problem and requires testing multiple hypotheses against a number of reasonable alternatives.[3]

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 numerous open-end mutual funds, are available for consideration by such a fiduciary. Two main issues plague the identification of the set of funds deemed 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.[4] 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. For example, the ability of a fund to outperform may be short-lived, as evidence of successful strategies attracts competitors (Hoberg, Kumar, and Prabhala, 2018) as well as additional investor flows.

A second major issue is whether an investor can, *ex-ante*, locate the number and identity of the set of active funds that will outperform in the future (with a high level of probability).

---

[1]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). 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.

[2]See, e.g., Cremers and Petajisto (2009), Kacperczyk, Sialm, and Zheng (2005, 2008), and Wei, Wermers, and Yao (2015).

[3]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.

[4]Recent studies suggest that the ability of individual mutual funds to outperform varies over time. For example, Kacperczyk, van Nieuwerburgh, and Veldkamp (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, Hollifield, Kacperczyk, and Kogan (2011) find evidence of strong (weak) predictability of mutual fund returns following periods of high (low) market returns. Ferson and Schadt (1996), Christopherson, Ferson, and Glassman (1998), Avramov and Wermers (2006), and Banegas, Gillen, Timmermann, and Wermers (2013) show that macroeconomic state variables can be used to better predict the future performance of individual equity mutual funds.

Estimates of funds' 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, Sialm, and Zheng, 2005). As a result, 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, van Nieuwerburgh, and Veldkamp (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 this task more challenging.[5]

Mutual funds' fleeting competitive advantage–and the ensuring challenge in identifying transient skills–has been noted and 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 in which fund alphas converge toward zero as a result of decreasing returns to scale and investors chasing past fund performance as a (noisy) signal of fund manager skills. Mamaysky, Spiegel, and Zhang (2008) develop a model in which managers observe private information signals which revert toward being uninformative. Similarly, Glode, Hollifield, Kacperczyk, and Kogan (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.

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 across funds to eliminate those funds whose performance is dominated by at least one other fund.[6] Since our stepwise approach for selecting a set of superior funds accounts for estimation error, investors can be confident that the funds that remain after the elimination process are most likely to genuinely have positive future alphas.

---

[5]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 exactly when investors more highly value significant active investment skills that may increase portfolio expected return without a commensurate increase in portfolio risk.

[6]As we will show, this pairwise comparison is efficient, in part, because it discards all funds deemed to be the "worst fund" in comparison with any of the other funds, eliminating one "worst fund" per round of comparisons. With traditional multiple hypothesis tests, the full (estimated) covariance matrix must be employed, which requires the estimation of the covariance of each fund with every other fund—a monumental task.

The sequence of pairwise tests seemingly ignores potential portfolio diversification benefits from considering funds with positive but slightly smaller alphas than those in the chosen set. Fund alphas are, however, not observed, and so investors must rely on estimated alphas when forming portfolios. This is important because alpha estimates tend to be estimated with considerable noise. This matters particularly for funds whose alpha estimates are positive but close to zero so that their true alphas are more likely to be negative or zero. Conversely, the true alphas of funds with large and positive t-statistics of alpha estimates are less likely to be zero or negative. We show that the tests we use in our pairwise elimination rule can be viewed as the information ratio of a long position in one fund and a short position in another, with positions chosen according to the predicted sign of the risk-adjusted return. The test statistic used by our approach will be high when the average risk-adjusted return on one fund is significantly higher than that of another fund and, also, when the two funds' returns are highly correlated. Both conditions indicate that the worse fund contributes little to the portfolio's information ratio. It follows that our stepwise approach can be seen as an elimination rule for deciding whether a fund is so unlikely to improve the information ratio of the overall portfolio, that it is tossed out. Thus, the FCS helps in controlling estimation error in the final portfolio formation process. Essentially, the FCS approach can be viewed as a first (selection) step used to identify those funds whose alphas we can most confidently trust are truly positive. From this perspective, the FCS approach can be viewed as a first step towards forming a portfolio of funds with positive alphas. The second step then applies mean-variance optimal weights to the funds included in the FCS.

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, 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.[7]

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.[8] 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,

---

[7]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.

[8]While the "False Discovery Rate" (FDR) 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, Scaillet, and Wermers (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%.

and Nason (2011), HLN henceforth, 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 pairwise tests to sequentially eliminate funds with inferior performance. 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.[9] 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).

The FCS is designed to choose the best among a set of funds using a particular forecasting model. The successful implementation of the FCS is therefore closely linked to how accurate the underlying performance model is in estimating and predicting fund performance. In particular, if the approach used to estimate fund performance is very noisy, the FCS is unlikely to have much power to discern differences between mutual funds and isolate the best funds. To overcome this issue, 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, Spiegel, and Zhang (2008); and (ii) extracts alpha estimates, pooling information from past fund returns and fund holdings. Empirically, we find that the FCS 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, a portfolio of funds included in the top FCS generates a four-factor alpha of more than 65 bps/month which is highly statistically significant and far higher than the performance achievable by a range of existing approaches.

The performance of most funds is estimated with large sampling errors. Our approach therefore works best if we use a relatively stringent criterion for inclusion of funds in the FCS–resulting in a relatively narrow set of superior funds 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 fraction of superior funds is larger in low-volatility states, and smaller in high-volatility states, suggesting that funds' ability to outperform is state-dependent.

We also find that dispersion in fund performance is greater during times with high levels of

---

[9]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.

market volatility, leading to the best funds–as identified by our FCS approach–performing better during such periods, relative to calmer times.[10] This is consistent with the model of Kacperczyk, van Nieuwerburgh, and Veldkamp (2016), which suggests that the rewards to skilled investors from processing information are higher when market risk is high.

The FCS methodology can also be used to 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). Funds identified as being inferior by our FCS go on to produce large negative and highly significant risk-adjusted returns, 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, Timmermann, Wermers, and White (2006) ask whether any fund is capable of genuinely outperforming on a risk-adjusted basis. 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, Scaillet, and Wermers (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. 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 refined set of superior funds is of great value to such investors. Further, Barras, Scaillet, and Wermers (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 Barras, Scaillet, and Wermers (2010).[11]

We show empirically that our approach, which endogenously builds the set of funds (starting with pairwise individual fund comparisons) with superior expected performance, is better at identi-

---

[10]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.

[11]Harvey and Lui (2018) propose a novel random effects approach that exploits cross-sectional information on alpha estimates in a way that reduces the effect of estimation error in individual fund alphas. While their approach is distinctly different from ours, like us they use information on the population of alpha estimates to identify funds with superior performance.

fying truly superior funds and forming portfolios of such funds. Compared to existing alternatives, our approach helps to discard individual funds that lie in the top past-performance fractile due to luck. Moreover, 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 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 FCS 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 concludes.

2. The Fund Confidence Set

This section introduces our approach to fund selection that endogenously determines the set of funds with superior performance, allowing the range of this set to vary over time. Our approach reflects how a sophisticated investor with access to historical data on individual fund returns and periodic security holdings could go about identifying superior funds, mimicking the search for skill described in papers such as Garleanu and Pedersen (2018).

Before formally introducing our Fund Confidence Set approach, we first provide a motivating example in which the set of funds is small enough that we can gain intuition for how the approach works. Next, we formally describe the approach, characterize its properties, and provide details on how we implement the approach on our mutual fund data.

2.1. Selecting funds with superior performance: A motivating example

To illustrate the central idea of our FCS approach and demonstrate how it differs from other possible approaches for fund selection, we initially consider an example with only four funds, namely Shearson Appreciation, Growth Fund of America, Sogen International Fund, and Pioneer II. We assume that an investor has obtained alpha estimates for these funds based on data available in November 1990 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 the statistical

6

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, Shearson Appreciation, Growth Fund of America, and Sogen International Fund 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.[12] In Step 1, all six pairwise t-statistics for the four funds are presented. The largest (in absolute terms) test statistic (-2.04; reported, in the table, for the funds in the rows relative to the funds in the columns) stems from comparing Sogen International Fund and Pioneer II.[13] Using a bootstrap methodology described below, we conclude that the performance of Pioneer II is significantly worse than that of Sogen International Fund, and we, therefore, eliminate Pioneer II.[14] In Step 2, we consider the three test statistics for pairwise comparisons of the remaining funds. The largest test statistic (2.03) arises from the comparison of Growth Fund of America and Shearson Appreciation. Using bootstrapped critical values, we can eliminate Shearson Appreciation. In Step 3 we only have one comparison left. The comparison of the Sogen International Fund and Growth Fund of America results in a test statistic of 0.29, which would not lead to elimination of any of the remaining funds at most conventional levels of significance.

Note from this example that Pioneer II, which has a positive but statistically insignificant alpha, could not be eliminated based on a pairwise comparison with Growth Fund of America alone. Instead this fund was eliminated through its pairwise comparison with Sogen International Fund. Likewise, Shearson Appreciation, which has a significantly positive alpha, could only be eliminated after a pairwise comparison with Growth Fund of America. The fact that the two

---

[12]This stepwise approach is similar to what we will be using in our subsequent analysis.

[13]The tests use the range statistic of Hansen, Lunde, and Nason (2011), which we discuss further below.

[14]A key contribution of Hansen, Lunde, and Nason (2011) is to calculate p-values for all the pairwise test statistics that remain valid in spite of the stepwise comparisons.

remaining funds each eliminate one of the other funds, but that neither of them can eliminate both, is explained by the differences in return correlations between the funds. Returns on Growth Fund of America are much higher correlated with returns on Shearson Appreciation than with returns on Sogen International Fund, while Sogen International Fund is much higher correlated with Pioneer II than with Growth Fund of America. 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.

The elimination step of the FCS uses a t-test that divides differences in funds' abnormal returns by the standard error of this difference, which tends to be lower, the higher the correlation between the funds' idiosyncratic returns. Hence, the ability of the FCS to eliminate inferiors fund will be higher, (i) the lower the funds' risk-adjusted returns relative to that of at least one other fund; and (ii) the more highly correlated the inferior funds' returns are with other fund returns. Conversely, the approach is less likely to eliminate inferior funds whose returns are only weakly correlated with that of other funds–unless, of course, their expected return is far smaller. From a portfolio perspective, it is desirable to eliminate low-alpha funds whose returns are highly correlated with those of at least one other fund while keeping funds whose returns are only weakly correlated with those of other funds since it preserves diversification gains.

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. Our objective is, therefore, to identify funds whose risk-adjusted performance can be reliably predicted.

Following common practice, we compute fund i's risk-adjusted return in period t by adjusting its excess returns, net of the T-bill rate, $R_{i,t}$, for its exposure to a set of risk factors, $\mathbf{z}_t$:

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

Here $\varepsilon_{i,t} \sim (0, \sigma^2_{\varepsilon_i})$ is the fund's idiosyncratic return, $\boldsymbol{\beta}_{i,t}$ measures the fund's exposure to the common risk factors, while $\alpha_{i,t}$ measures its risk-adjusted (abnormal) performance at time t. The model in equation (1) is quite general as it allows both α_{it} and $\boldsymbol{\beta}_{it}$ 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.

Let $\hat{\alpha}_{i,t|t-1}$ be the expected value of fund i's alpha during period t based on information available at time $t-1$. To determine if $\hat{\alpha}_{i,t|t-1}$ reliably predicts fund i's performance in period t, consider

the following "predictive alpha":

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

Here the sign function $sign(\bullet)$ equals $+1$ if the argument is positive, -1 if the argument is negative, or 0 if the argument is zero. $P_{i,t}$ is the risk-adjusted return an investor would earn at time t from using $\hat{\alpha}_{i,t|t-1}$ as a buy or sell signal at time $t-1$. $P_{i,t}$ will be large for fund i if either (i) a positive risk-adjusted return $(R_{i,t} - \boldsymbol{\beta}'_{i,t}\mathbf{z}_t > 0)$ is anticipated by a positive alpha estimate $(\hat{\alpha}_{i,t|t-1} > 0)$, or (ii) a negative risk-adjusted return $(R_{i,t} - \boldsymbol{\beta}'_{i,t}\mathbf{z}_t < 0)$ is anticipated by a negative alpha estimate $(\hat{\alpha}_{i,t|t-1} < 0)$. Conversely, the objective in (2) penalizes funds with positive alpha forecasts but negative future risk-adjusted returns, or vice versa. Predictive alpha measures the risk-adjusted return an investor would have earned from following the "directional" information in the alpha forecast, buying if the alpha forecast is positive, otherwise selling. It preserves scale as it is measured in return units and rewards (penalizes) alpha forecasts associated with higher returns (larger losses) for the investor.[15]

The predictive alpha measure can also be interpreted as the realized alpha, conditional on the sign of its predicted value. The measure is symmetric as it rewards correct forecasts of both positive and negative excess returns. When we apply the measure to select funds with superior performance, we therefore limit our attention to funds whose future predicted alphas are positive.

We estimate fund i's predictive alpha at time t using the sample average of (2):

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

Here, $\hat{\alpha}_{i,\tau|\tau-1}$ is the forecast of $\alpha_{i,\tau}$ based on information available at time $\tau - 1$, and $\hat{\boldsymbol{\beta}}'_{i,\tau}$ are least-squares estimates of $\boldsymbol{\beta}'_i$, using data only up to time τ. Finally, $t_{i0} + 1$ is the starting point of the sample used to estimate the ith fund's predictive alpha performance up to time t, $\bar{P}_{i,t}$. We estimate the average predictive alpha on at least 12 monthly return observations, but allow up to 60 observations to be included in the estimate if available.

2.3. Fund Confidence Set

Our data contain almost 3,000 funds whose performance needs to be pairwise 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. This approach is designed to choose the set of "best" forecasting models from a larger set of candidate models and so needs to be modified

---

[15]Appendix A considers a range of alternative specifications for the predictive alpha measure.

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, with a certain probability, 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 $\mathcal{F}_t^0 = \{F_{1t}, ..., F_{nt}\}$ be the initial set of funds under consideration while $P_{i,t}$, 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

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

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 fund deemed to be worst relative to another fund in the current set of surviving funds, $\mathcal{F}_t \subseteq \mathcal{F}_t^0$–provided that this difference is statistically significant. Each round of this procedure tests the null hypothesis of equal performance

$$H_{0,\mathcal{F}_t} : \mu_{ij} = 0, \text{ for all } i,j \in \mathcal{F}_t, \tag{5}$$

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

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

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 λ 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 $H_{0,\mathcal{F}_t} : \mu_{ij} = 0$ for all $i,j \in \mathcal{F}_t^0$ at a critical level λ. If $H_{0,\mathcal{F}_t}$ is accepted, the FCS is $\hat{\mathcal{F}}_t^* = \mathcal{F}_t$. 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 $\hat{\mathcal{F}}_t^*$.

To choose whether or not to eliminate individual funds, we use equations (3) and (4) and estimate the performance of fund i relative to fund j as $\bar{d}_{ij} = t^{-1} \sum_{\tau=1}^{t} d_{ij,\tau}$. Next, we divide this

measure by its standard error, $\sqrt{\widehat{var}(\bar{d}_{ij})}$, to obtain a t-test[16]

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}}. \tag{7}$$

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

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

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 joint-normally distributed with unknown covariance matrix, Ω. Because so many pairwise test statistics are being compared, 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:[17]

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

Based on (8) and (9), we purge any fund whose performance looks sufficiently poor, relative to that of at least one other fund currently included in the FCS.[18]

Carrying out a large number of possibly dependent tests makes it far from trivial to control the coverage probability of this stepwise procedure. Notably, if each round conducts a test at a fixed critical level, λ, 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 coverage probability, $1 - \lambda$, of the FCS by dynamically adjusting the bootstrapped critical values employed in each elimination step.[19]

Funds may be included in the FCS because they have a precisely estimated positive average performance. Indeed, when the performance data allow for sharp inference, the equivalence tests first eliminate the poor funds before questioning the superior funds. However, funds with negative

---

[16]Dividing $\bar{d}_{ij}$ by its standard error gives us a pivotal test statistic with better sampling properties.

[17]Our objective is to eliminate funds for whom our risk model has poor forecasting power, while HLN's objective is to eliminate the least accurate forecasting models. Accordingly, we alter the elimination rule, as compared to HLN, who use $\arg \max_{i \in \mathcal{F}_t} \sup_{j \in \mathcal{F}_t} t_{ij}$.

[18]Specifically, if the FCS p-value for the fund identified by (8) is smaller than the λ-quantile of the bootstrapped distribution, then this fund is deemed inferior to at least one other fund and is eliminated.

[19]Theorem 1 in Hansen, Lunde, and Nason (2011) establishes conditions under which the probability that a truly superior fund is contained in the estimated FCS is greater than or equal to $1 - \lambda$, while the probability of wrongly including an inferior fund in $\hat{\mathcal{F}}_t^*$ asymptotically goes to zero.

alphas may fail to be eliminated from the FCS because their alphas are estimated with large standard errors, which reduces t_{ij} in (7) and makes it hard to reject the null. When the performance data does not support sharp inference, the FCS algorithm will tend to eliminate too few funds, resulting in an over-sized set that includes many inferior funds. The trade-off between these effects depends on the choice of λ as we next discuss.

2.4. Choosing λ

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 any 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 is controlled by the choice of the size of the equivalence test (λ) used by the FCS approach, which, therefore, is an important parameter.

Setting λ 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 λ 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.

In practice, the alpha estimates of many of the funds in our sample are quite noisy and so we choose relatively high values of λ. Our choice reflects that returns data often do not have much power to distinguish between the performance of different funds.[20]

In trading off these effects, our goal is to identify a set of funds whose expected risk-adjusted performance is genuinely superior. 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.[21]

We, therefore, opt for a relatively high value of λ, choosing $\lambda = 0.90$ as our benchmark. However, to illustrate how sensitive our results are to our particular choice of λ, in robustness tests, we also 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 λ-values as tight ($\lambda = 0.90$), medium-tight ($\lambda = 0.75$), medium ($\lambda = 0.50$), and wide ($\lambda = 0.10$).

---

[20]See also Harvey and Lui (2018) for a thorough discussion of the trade-offs between type I and type II errors in financial decisions.

[21]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).

2.5. Candidate set of funds

The MCS approach requires conducting all possible pairwise comparisons to determine if any model is inferior and, if so, which model gets eliminated. A key contribution of HLN is to show that this exhaustive protocol for testing and comparing models controls the size of the resulting step-wise approach. This assures us that the FCS approach has well-defined properties for identifying superior funds and eliminating inferior ones.

With almost 3,000 funds in our sample, it is, however, not feasible to conduct all possible pairwise performance comparisons.[22] One could consider alternative approaches to reducing the number of pairwise comparisons such as picking the fund with the highest estimated $\bar{P}_{it}$ ("top fund") and comparing other funds against this.[23] This approach is likely to eliminate a large number of funds without requiring nearly as many pairwise comparisons. Relative to this simple and intuitive approach, the stepwise FCS approach seems quite inefficient. A limitation of the "top fund" approach is that its theoretical properties (e.g., controlling the size of the test) have not been established and so it is unclear how well it would perform in practice. In Appendix B, we therefore conduct a Monte Carlo simulation study that explores the properties of this simpler approach. We find that the level of the associated sequence of tests cannot be controlled by this simple algorithm. In particular, the probability of correctly including all superior funds in the set drops towards zero as the number of superior funds increases. Comparing other funds' alpha estimates only to the top-ranked fund disregards that this fund's performance is, itself, the result of random sampling variation. Intuitively, the alpha estimate of the top-ranked fund is more likely to be biased upwards, as a result of random sampling variation, than that of a fund ranked in the middle. The "top fund" approach is, therefore, likely to eliminate too many funds from the equivalence (FCS) set. Conversely, by conducting all pairwise comparisons, including seemingly redundant ones, the FCS approach avoids anchoring the performance on a single fund's alpha estimate and, thus, properly accounts for the effect of sampling variation on the funds' alpha estimates.

Given these findings, as a simple way to reduce the number of comparisons, we only consider funds with positive estimates of predictive alpha in the initial set of funds ($\mathcal{F}_t^0$). Heuristically, this can be thought of as a simple nonparametric way to eliminate funds with poor predictive alphas.

2.6. Summary of the FCS procedure

We briefly summarize the different steps that make up the FCS algorithm:

---

[22]For example, a cross-section of 3,000 funds observed over 313 months requires over 1.4 billion pairwise comparisons. We estimate the fund confidence set using the MulCom 2.0 package for Ox; see Hansen and Lunde (2010) and Doornik (2006).

[23]We are grateful to a referee for suggesting this alternative method.

Step 1: For all funds, i, satisfying the minimal data requirements, compute alpha forecasts for period $t + 1$, $\hat{\alpha}_{i,t+1|t}$.

Step 2: Discard funds with negative or zero alpha forecasts, $\hat{\alpha}_{i,t+1|t} \leq 0$.

Step 3: Evaluate the predictive alpha, $P_{i,t}$ for the remaining funds.

Step 4: Select funds with positive estimates, $\bar{P}_{i,t}$.

Step 5: Estimate the FCS based on the candidate set of funds from Step 4.

Step 6: Construct a portfolio based on the funds included in the FCS.

Step 7: Obtain the return for this portfolio.

We repeat these steps at each point in time, t, during our sample.

2.7. FCS and portfolio formation

Assuming that returns are generated by a factor model, Treynor and Black (1973) show that mean-variance efficient portfolios of risky assets weight individual assets in proportion with their alphas, α_i, divided by the variance of their idiosyncratic risk, σ_i^2, relative to that of all included funds. Specifically, for funds $i = 1, ..., n$ with positive alphas, $\alpha_i > 0$, the optimal weights, ω_i^*, are

$$\omega_i^* \propto \frac{\alpha_i}{\sigma_i^2} = \frac{IR_i}{\sigma_i}, \tag{10}$$

so the optimal portfolio weights are proportional to the information ratio, IR_i, scaled by the idiosyncratic standard deviation.

Portfolio effects are seemingly ignored by the FCS approach which considers individual funds' performance without accounting for possible diversification benefits from including funds with small, but positive, alphas. However, the weights in (10) are idealized and, in practice, investors must rely on estimated alphas when forming portfolios. This matters particularly for funds whose alpha estimates are positive but close to zero so that their true alphas are more likely to be negative. Conversely, it is less likely that the true alphas are negative for funds with large and positive alpha t-statistics.

The t-test we use in our pairwise elimination rule in equation (7) can be viewed as the information ratio of a long position in one fund (fund i) and a short position in another (fund j), with positions chosen according to the sign of the predictive alpha. If the test statistic is high, either the risk-adjusted return on fund i is high compared to that of fund j or the two are highly correlated, or both. The low return of fund j and/or the high correlation will indicate that fund j contributes little to the mean-variance efficient portfolio's information ratio. It follows that the FCS can be seen as an elimination rule that decides whether a fund is so unlikely to improve the information ratio of the overall portfolio, that it is tossed out. Thus, the FCS helps in controlling estimation

error in the portfolio formation process and it can be viewed as a natural first step in forming a portfolio of funds with positive alpha estimates that, moreover, accounts for cross-fund correlation in residual risk. The second step then applies mean-variance optimal weights to the funds selected in the first-step FCS, assigning greater weight to funds with large IR values per unit of idiosyncratic risk.

To directly assess these arguments and evaluate the practical usefulness of the FCS approach in a portfolio context, we conduct three experiments. First, we compare the performance of the FCS portfolio to that of a portfolio applying mean-variance optimal weights to all funds with positive alpha estimates. If the FCS fund performs better than this broader portfolio, this suggests that using the FCS as a first step to select those funds most likely to have positive alpha values works well.

Second, comparing the performance of portfolios of FCS-selected funds that apply (i) optimal weights from (10) to (ii) equal weights, we can see if differences in scaled IRs (IR_i/σ_i) across funds included in the FCS are large enough to be exploited to improve portfolio performance. If the differences in scaled IRs are sufficiently large and can be estimated sufficiently precisely, we would expect the optimally-weighted portfolio to produce notably larger gains than the equal-weighted portfolio.

Third, many investors have constraints on the number of funds they can invest in. This could force them to either select a smaller subset of the funds in the FCS (in periods where the FCS contains many funds) or, conversely, to add more funds (in periods where the FCS contains very few funds). We therefore also consider portfolios that are subject to holding a minimum number of funds (e.g., $n \geq 10$) or a maximum number of funds (e.g., $n \leq 10$).

In each case we use information from the fund elimination process adopted by the FCS approach to construct a list of funds investors can choose from. Because our approach is structured around a number of equivalence tests, if the FCS list is too long, without loss of generality, investors can simply choose a subset of the funds selected. Conversely, if the list is too short, they could include the last funds to be eliminated from the FCS.

3. Performance measurement model

This section introduces our approach to estimate the risk-adjusted performance of individual funds. The performance measurement model plays an important role in the empirical analysis as no approach can be expected to work well without accurate signals about individual funds' future performance. As we show below, an approach that uses both returns and fund holdings data can provide sharper inference about alphas than an approach which uses only returns data. This, in turn, 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 as our benchmark a four-factor model that, in addition to the market factor, adjusts for the size and value factors of Fama and French (1993) and the momentum factor of Carhart (1997): The benchmark four-factor model takes the form

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i' \mathbf{z}_t + \varepsilon_{it}. \tag{11}$$

Here R_{it} is the monthly return on fund i, measured in excess of a 1-month T-bill rate, and $\mathbf{z}_t = (R_{mt}, SMB_t, HML_t, MOM_t)'$ 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 (1993), 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.

Following common practice, we allow for slowly-evolving shifts in mutual fund performance and risk exposures by estimating $\boldsymbol{\theta}_i = (\alpha_i \ \boldsymbol{\beta}_i')'$ on a rolling 60-month data window.

3.1. Time-varying skills

We generalize (11) in two ways. First, following Mamaysky, Spiegel, and Zhang (2007, 2008), we assume that managers receive information (unobserved to the econometrician) that is correlated with future returns and gives rise to a time-varying component in fund performance. Second, we introduce a generalized framework that combines information from past returns with holdings data to more accurately extract an estimate of fund performance.

Individual funds' alphas are value-weighted averages of their stock-level alphas, so our analysis starts by decomposing the excess return of each stock, r_{jt}, into a risk-adjusted return component, α_{jt}, a systematic return component $\boldsymbol{\beta}_j' \mathbf{z}_t$, and an idiosyncratic return component, ε_{jt}.[24] We stack these return components into $N_t \times 1$ vectors $\boldsymbol{\alpha}_t$ and $\boldsymbol{\varepsilon}_t$ and an $N_t \times 4$ matrix of betas, $\boldsymbol{\beta}$, where N_t is the number of stocks in existence at time t. Because abnormal returns are likely to be temporary, we allow individual stocks' alphas to vary over time.

Fund returns can be computed by summing across individual stock returns, $\mathbf{r}_t$, weighted by the funds' ex-ante portfolio weights measured at the end of time $t-1$, $\boldsymbol{\omega}_{t-1}$. The excess return of fund i, net of the risk-free rate, transaction costs, fees and expenses, k_i, is given by

$$R_{it} = \boldsymbol{\omega}_{it-1}'(\boldsymbol{\alpha}_t + \boldsymbol{\beta}\mathbf{z}_t + \boldsymbol{\varepsilon}_t) - k_i \equiv \alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{z}_t + \varepsilon_{it}, \quad \varepsilon_{it} \sim iidN(0, \sigma_\varepsilon^2), \tag{12}$$

where $\alpha_{it} = \boldsymbol{\omega}_{it-1}'\boldsymbol{\alpha}_t - k_i$, $\boldsymbol{\beta}_{it} = \boldsymbol{\omega}_{it-1}'\boldsymbol{\beta}$, and $\varepsilon_{it} = \boldsymbol{\omega}_{it-1}'\boldsymbol{\varepsilon}_t$.

---

[24]Following Mamaysky, Spiegel, and Zhang (2008), we assume that stock betas are constant, although this assumption can be relaxed. Note that although individual stock betas are assumed constant, fund betas will be time-varying because of changes in the portfolio weights.

Following Mamaysky, Spiegel, and Zhang (2008), we capture individual funds' private information through a private (fund-specific) signal, F_{it}, which follows an autoregressive process,

$$F_{it} = \nu_i F_{it-1} + \eta_{it}, \quad \eta_{it} \sim iidN(0, \sigma_\eta^2), \text{ for } \nu_i \in [0; 1), \tag{13}$$

Following Mamaysky, Spiegel, and Zhang (2008), assume that (i) fund portfolio weights are linear in the private signal, $\boldsymbol{\omega}_{it-1} = \bar{\boldsymbol{\omega}}_i + \boldsymbol{\gamma}_i F_{it-1}$; (ii) $Cov(\varepsilon_{it}, \eta_{it}) = 0$; and (iii) there is a constant mapping $\bar{\boldsymbol{\alpha}}_i$ from fund i's private signal to its ability to identify abnormal performance, $\boldsymbol{\alpha}_{it} = \bar{\boldsymbol{\alpha}}_i F_{it-1}$. These assumptions and equation (12) imply that fund i's beta and alpha are linear and quadratic functions of the signal, F_{it-1}, respectively:

$$\alpha_{it} = \bar{\boldsymbol{\omega}}_i' \bar{\boldsymbol{\alpha}}_i F_{it-1} + \boldsymbol{\gamma}_i' \bar{\boldsymbol{\alpha}}_i F_{it-1}^2 - k_i \equiv a_i F_{it-1} + b_i F_{it-1}^2 - k_i, \tag{14}$$

$$\boldsymbol{\beta}_{it} = \bar{\boldsymbol{\omega}}_i' \boldsymbol{\beta} + \boldsymbol{\gamma}_i' \boldsymbol{\beta} F_{it-1} \equiv \bar{\boldsymbol{\beta}}_i + \mathbf{c}_i F_{it-1}. \tag{15}$$

Individual funds' signals, F_{it}, are unobserved to the econometrician who can, however, extract an estimate of the signal from fund returns by writing (13)-(15) in state space form:

$$R_{it} = a_i F_{it-1} + b_i F_{it-1}^2 - k_i + (\bar{\boldsymbol{\beta}}_i' + \mathbf{c}_i' F_{it-1})\mathbf{z}_t + \varepsilon_{it} \tag{16}$$

$$F_{it} = \nu_i F_{it-1} + \eta_{it}.$$

As explained by Mamaysky, Spiegel, and Zhang (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 (16).[25]

3.1.1. Information in fund holdings

Conventional approaches to estimating fund performance use data on past returns, which can be very noisy and, thus, reduces our ability to identify funds with superior performance. We therefore augment our model with information from portfolio holdings which, as is clear from equation (14), can be used to track how a fund's alpha evolves through time. Building on this idea, we next generalize the methodology in Mamaysky, Spiegel, and Zhang (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 in (16). 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 sensitivity to book-to-market, market

---

[25]After normalizing one of the elements, the four-factor model requires 13 parameters to be estimated.

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 stocks' abnormal returns across all stock positions held by a fund, we obtain the fund-level characteristic-selectivity (CS) measure of Daniel, Grinblatt, Titman, and Wermers (1997):

$$CS_{it} = \boldsymbol{\omega}'_{it-1}(\mathbf{r}_t - \mathbf{r}_{bt}). \tag{17}$$

Here, $\mathbf{r}_t$ and $\mathbf{r}_{bt}$ are vectors of excess returns on the stocks held by the fund ($\mathbf{r}_t$) and the corresponding 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 characteristics-matched stocks are chosen mechanically, the average stock can be expected to have zero alpha, $\boldsymbol{\alpha}_{bt} = \mathbf{0}$. Using (14) and (17) we, therefore, have

$$
\begin{aligned}
CS_{it} &= \boldsymbol{\omega}'_{it-1}\left(\boldsymbol{\alpha}_t + \boldsymbol{\beta}'\mathbf{z}_t + \boldsymbol{\varepsilon}_t - \left(\boldsymbol{\alpha}_{bt} + \boldsymbol{\beta}'_b\mathbf{z}_t + \boldsymbol{\varepsilon}_{bt}\right)\right) \\
&= \boldsymbol{\omega}'_{it-1}(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{bt}) + \boldsymbol{\omega}'_{it-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_b)'\mathbf{z}_t + \boldsymbol{\omega}'_{it-1}(\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_{bt}) \\
&= \alpha_{it} + k_i - \boldsymbol{\omega}'_{it-1}\boldsymbol{\alpha}_{bt} + (\boldsymbol{\beta}_{it} - \boldsymbol{\beta}_{bt})'\mathbf{z}_t + \varepsilon_{it} - \varepsilon_{bt} \\
&= a_i F_{it-1} + b_i F_{it-1}^2 + \varepsilon_{it} - \varepsilon_{bt}.
\end{aligned}
\tag{18}
$$

Because the CS measure does not depend on sample estimates of risk factor loadings, it has the potential to generate a more accurate estimate of fund performance and, thus, to improve inference based on the return-only models.[26]

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

$$
\begin{pmatrix} R_{it} \\ CS_{it} \end{pmatrix} = \begin{pmatrix} a_i F_{it-1} + b_i F_{it-1}^2 - k_i \\ a_i F_{it-1} + b_i F_{it-1}^2 \end{pmatrix} + \begin{pmatrix} \bar{\boldsymbol{\beta}}_i + \mathbf{c}_i F_{it-1} \\ 0 \end{pmatrix} \mathbf{z}_t + \begin{pmatrix} \varepsilon_{it} \\ \varepsilon_{it} - \varepsilon_{bt} \end{pmatrix} \tag{19}
$$
$$F_{it} = \nu_i F_{it-1} + \eta_{it}.$$

Compared to the model of Mamaysky, Spiegel, and Zhang (2008) in (16), our model uses the additional information contained in CS_{it}. This has the potential to make estimation and extraction of funds' private information component, F_{it}, more precise. We estimate the parameters of (19) using the extended Kalman filter with two measurement equations; Appendix C provides further details.

---

[26]The CS measurement equation is "idealized," in the sense that the CS performance measure will not capture intra-quarter trading returns. The CS and return-based measures can be viewed as different estimates of the same underlying fund performance, observed with different measurement errors.

4. Performance results

This section introduces our data and analyzes the empirical performance of FCS portfolios.

4.1. Data

Our empirical analysis uses monthly returns on a sample of U.S. equity mutual funds over the 32-year period from 1980:07 to 2012:12. Individual fund returns are taken from the Center for Research in Security Prices (CRSP) mutual fund data base, and are computed net of transaction costs and fees.

We use quarterly holdings data from Thomson Financial to construct a three-month estimate of CS.[27] We merge the returns and holdings data using the MFLINKS files of Wermers (2000), updated by the Wharton Research Data Services, which allows us to map the Thomson holdings data to CRSP returns using the funds' WFICN identifier. To reduce estimation error, we require each fund to have at least six months of contiguous returns data and 72 months of concurrent observations on returns and CS (not necessarily contiguous). In total, we have returns and holdings data on 2,991 actively managed U.S. domiciled equity mutual funds.[28] The number of funds included in the analysis peaks at above 2,500 in 2005 before declining to around 1,800 in 2012.

We use time-series data on individual funds' historical returns to obtain alpha estimates. Table 2 provides summary statistics for the cross-sectional distribution of individual fund (ex-post) alphas. Consistent with previous academic studies (e.g., Carhart, 1997), the median fund has a negative alpha–minus 78 bps/year for our sample. The bottom 5% of funds, ranked by alpha performance, have a negative alpha estimate of about -44 bps/month, or just under -5.2%/year. The top 5% of funds have an alpha estimate of 26 bps/month, or 3%/year. The cross-sectional distributions of Sharpe and information ratios for the funds (shown in the second and third columns) also suggest that only a small subset of funds is capable of outperforming.[29]

4.2. Performance of alpha-sorted portfolios

To establish a reference point for our FCS results, we consider both portfolio sorting alternatives from the existing literature as well as new strategies chosen for their comparability with our FCS approach in at least one dimension, e.g., the average number of funds held.

---

[27]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.

[28]Sector funds, defined as funds whose R^2 is less than 70% in a four-factor regression, are excluded because the simple four-factor risk-adjustment approach is not appropriate for these funds.

[29]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, Spiegel, and Zhang (2008).

First, we consider a portfolio of all funds with positive alpha forecasts. Second, to more directly evaluate the effect of applying the FCS approach for fund selection, we consider a portfolio of the same funds from which the FCS is determined, i.e., funds with positive alpha forecasts and positive average predictive alphas. Third, we consider subsets of the candidate set of funds which each month rank funds on their alphas $\hat{\alpha}_{i,t+1|t}$ and form optimally-weighted portfolios.

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 optimally-weighted portfolios based on the funds' updated alpha estimates, then, again, record their next-month returns. Using five years of data to initiate the portfolio sorts and another year to obtain an estimate of the predictive alpha, we get a time series of portfolio excess returns, R_{pt}, spanning the 26-year period from July 1, 1986 to December 31, 2012. This approach, while requiring 6 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.[30]

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},

$$R_{pt} = \alpha_p + \boldsymbol{\beta}_p' \mathbf{z}_t + \varepsilon_{pt}, \quad t = 1, \dots, T. \tag{20}$$

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

Table 3 presents performance estimates for the alpha- and predictive alpha sorts based on three different alpha models, namely the latent skill holdings (LSH, top panel), latent skill (LS, middle panel) and the conventional constant-alpha Carhart model (bottom panel). The table shows results for both optimally- and equal-weighted portfolios.

We find notably weaker performance of the alpha-sorts applied to the top-ranked funds compared to the predictive-alpha sorts for the latent skill and latent skill holdings models that are set up to capture time-variation in the underlying alphas. For example, under the LS model, the alpha of the optimally weighted portfolio consisting of the top 5% of funds goes from 2 bps/month to 33 bps/month once we move from ranking on alphas to ranking on predictive alphas. For the LSH model we see a rise from 2 bps/month to 25 bps/month.

In turn, for the Carhart alpha estimates, there is little difference between the performance results that use the regular alpha measure (left columns) and the results based on our predictive alpha measure (right columns).[31] If alphas are genuinely constant (as assumed by the conventional Carhart specification), the predictive alpha measure simply adds a bit of noise from the idiosyncratic shocks. Hence, we should expect the differences in the performances of the alpha and predictive

---

[30]See also Carhart (1997) for a discussion of survivorship bias.

[31]Our estimates of the top-5-10% ranked funds' Carhart alphas are very similar to those reported in Mamaysky, Spiegel, and Zhang (2007) for a portfolio of similar size.

alpha specifications to be quite small and perhaps a little weaker for the predictive alpha measure. This is exactly what we find.

We conclude from this evidence that using our proposed predictive alpha approach can help improve the empirical performance of portfolios constructed using models that allow for time-varying alphas. However, the empirical performance estimates remain relatively weak. As we next show, much stronger results can be obtained by relaxing the assumption that the proportion of funds used to form portfolios is fixed through time and across economic states.

4.3. Performance of superior funds

Figure 1 illustrates how, at a single point in time (November 1994), the FCS approach selects superior funds from the larger set of funds with positive forecasts of alpha. In each case, distributions refer to average predictive alphas. Gray bars plot the distribution for the full set of eligible funds in existence at that point. This distribution is centered close to zero, and has a wide dispersion. We also show distributions for funds with positive forecasts of alphas (green) for next month (December 1994) and for funds with positive alpha forecasts as well as positive estimates of average predictive alpha (red). Finally, the blue bars capture the distribution for funds included in the FCS. Note that the FCS distribution lies to the right of the distribution for the more inclusive candidate set of funds with positive average predictive alphas and the FCS methodology is markedly more discriminating than an approach that simply selects funds with positive alphas.

Turning to the formal analysis, the left columns in Table 4 present performance measures for portfolios formed each month by applying either optimal- (top four rows), equal- (middle four rows), or winsorized optimal-weights to the set of funds included in the FCS. As in Table 3, the sample runs from 1986:07 to 2012:12. We consider a range of values for the size of the equivalence tests, $\lambda = \{0.90, 0.75, 0.50, 0.10\}$, corresponding to tight, medium-tight, medium, and wide fund confidence sets.

First, consider the results for the latent skill holdings approach (top panel). As we move from $\lambda = 0.10$ to $\lambda = 0.90$, we observe a monotonically increasing pattern in alpha performance, rising from minus 1 bps/month for the wide FCS to 66 bps/month for the tight FCS, the latter being highly statistically significant with a t-statistic of 3.25.

These findings suggest that the marginal funds included in the wider FCS perform worse than the funds identified by the more selective approach. Because the alpha estimates of many funds are surrounded by large standard errors, choosing a small value of λ means that the equivalence test used to eliminate inferior funds has insufficient power to reject the null that fund alphas are identical; we analyze this finding in more detail below.

The sequential elimination of funds from the candidate set, Step 5 in Section 2.6, is important to the FCS results. In particular, the alpha estimate of a portfolio consisting of the entire candidate

set from which our FCS is determined, is zero bps/month–much lower than the alpha of the tight FCS specification. For comparison, we can consider a portfolio composed of the same average proportion of funds from the candidate set ranked on average predictive alpha. This portfolio, constructed to have the same average "breadth" as the tight FCS generates a performance of 13 bps/month–much lower than the tight FCS portfolio even though the number of funds is the same. Selecting a narrow set of funds is, therefore, not sufficient to achieve superior performance.

Next, consider the results for the latent skill and constant alpha (Carhart) models. Using equally-weighted portfolios, these approaches generate FCS-tight alphas of 20 and 34 bps/month which is economically sizeable, though not statistically significant and notably lower than the values produced by the tight FCS portfolio constructed from the latent skill holdings model (67 bps/month with a t-statistic above three). Using portfolio weights chosen to optimize the information ratio, we find similar performance results for the Carhart alphas, though somewhat weaker performance for the latent skill model. This deterioration is again consistent with Mamaysky, Spiegel, and Zhang (2007) who find that an unfiltered latent skill model can get dominated by outliers which is the reason they introduce two filters on the alpha estimates to obtain more stable results. Consistent with their finding, when we winsorize the portfolio weights to limit the influence of the funds with the ten percent highest information ratios, the performance results improve substantially for the latent skill model. Conversely, winsorization has little effect on the portfolios constructed from the LSH model, strengthening the case for using holdings information to obtain alpha estimates that are less sensitive to outliers.

Overall, the improvements in alpha estimates observed as we move from the FCS-wide to the FCS-tight portfolio are smaller for the constant-alpha Carhart model (an increase from 3 bps/month to 32 bps/month) than what we observe for our more accurate latent skill holdings model (an increase from 4 bps/month to 67 bps/month). Using more accurate alpha estimates with stronger predictive power over future fund returns thus clearly helps the performance of the FCS approach.

These observations suggest that the large alpha values of our tight FCS approach result from the joint effect of using a more accurate approach to estimating alphas (pooling return and holdings data) as well as using the sequential FCS fund selection approach to eliminate inferior funds. Interestingly, the performance of the FCS portfolios is robust to whether these portfolios are constructed using weights proportional to funds' IRs or using equal weights. This confirms the importance of the selection step conducted by the FCS–choosing the right set of superior funds is more important than how they are weighted in a portfolio going forward.

Table 4 also reports Sharpe ratios and information ratios for the optimally-weighted and equal-weighted portfolios. Before interpreting these measures, note that the Sharpe ratios of the portfolios that choose weights to maximize the information ratio need not exceed the Sharpe ratios of the equal-weighted portfolios. Conversely, in the absence of estimation error, we would expect the information ratios of the IR-weighted portfolios to be higher than the information ratios of the

equal-weighted portfolios. In general, we do not find that this holds – in fact, the Sharpe ratios and information ratios of the portfolios weighted to maximize the IR tend to be slightly smaller than the information ratios of the equal-weighted portfolios.

To shed light on whether this finding is due to estimation errors in alphas and idiosyncratic risk levels, we also consider a weighting scheme that winsorizes the IR-based portfolio weights, thus reducing the weights on funds with the very largest IR estimates as explained in Appendix A. These weights are less sensitive to estimation error than the weights that are proportional to the funds' estimated information ratios. Consistent with estimation error in information ratios being the chief reason for why the IR-weighted portfolios fail to increase the information ratio of the equal-weighted portfolios, we find that the winsorized IR-weighted portfolios tend to generate higher information ratios than the portfolios that use weights that are proportional to the estimated IRs. As a final data point consistent with this explanation, the average ex-ante IR-values of the optimally-weighted, optimally IR-weighted and winsorized, and equal-weighted FCS-tight portfolios are 1.30, 1.15, and 0.78, whereas the actual (ex-post) IR values of these three portfolios are 0.74, 0.76, and 0.77. This shows that the information ratio of the optimally weighted portfolio is more sensitive to estimation error (i.e., sees a larger percentage decline from ex-ante to ex-post IRs) than the winsorized and equal-weighted portfolios.

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 only to redefine the candidate set of funds from Step 4 to consider funds with negative alpha forecasts and positive average predictive alpha–the latter to ensure that a fund's performance is actually predictable.

The right columns in Table 4 consider the performance of a strategy that identifies funds with negative alpha forecasts and positive average predictive alphas, followed by the use of the FCS approach to select the set of worst-performing funds. Again, we start with the results for the latent skill holdings model. The tight FCS portfolio generates the smallest alpha estimates–around -34 bps/month–while the medium and wide FCS portfolios generate alphas around -15 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, Spiegel, and Zhang (2008) model that does not use holdings information, although the results are a little less sharp for the constant-alpha approach. Because the left tail of the performance distribution is thicker than the right tail, the performance of the portfolios of inferior funds is less sensitive to our choice of λ than portfolios of superior funds. For inferior funds, the main effect of varying λ is to include or exclude funds with slightly smaller negative performance than the worst-performing funds.

4.5. Time variation in fund performance

The values reported in Tables 3 and 4 are estimates of average alpha performance and thus 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 D 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 tight FCS portfolio (in blue), along with the performance of a portfolio of the top 5% predictive alpha-ranked funds in the candidate set (in purple) from which we select the FCS. The red line tracks the performance of the candidate set of funds. Finally, the green line tracks the performance of the set of funds with positive alpha forecasts, while the black line tracks the performance of the average fund at a given point in time. Each portfolio uses weights that are proportional to the individual funds' information ratios in accordance with equation (10).

In the first part of the sample (until 1994), the local alpha estimates of the FCS portfolio fluctuate around zero. Starting in 1994, the performance of the FCS portfolio increases markedly, before peaking at about 150 bps/month in 1999-2001. Over the same five-year period the top 5% of funds from the candidate set also performs well, though with alphas that clearly are lower than those of the FCS. The performance of the FCS portfolio decreases to around 40 bps/month in 2003, fluctuates between 40 and 60 bps/month for until 2010 before briefly decreasing to zero in 2011 and bouncing back to 40 bps/month at the end of the sample. The increase in the performance of funds included in the FCS towards the end of the sample is particularly noteworthy given the evidence of intensified competition in the fund industry (Pastor, Stambaugh, and Taylor, 2015).

Figure 3 plots the time-series evolution in the local alpha estimates for the tight FCS portfolio of inferior funds. This portfolio succeeds in generating negative risk-adjusted returns in the entire sample. For most of the sample, the inferior FCS funds generate negative returns between -40 and -20 bps/month. However, the performance of the FCS portfolio of inferior funds is particularly poor between 1996 and 2000 and at times falls below 70 bps/month.

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 the evaluation of the FCS portfolio.[32] Funds selected by the tight FCS ($\lambda = 0.9$) go on to produce five-factor alpha estimates of 47 bps/month with a t-stat of 2.36, suggesting that our performance results are quite robust to including a technology factor.

4.7. Alternative portfolio approaches

Earlier, we argued that the FCS can be viewed as a first step for determining the set of funds whose alphas are most likely to truly be positive. Having identified such funds, this step can be followed by a portfolio formation step which assigns weights on each fund so as to maximize the overall information ratio.

Because a information-ratio maximizing investors would want to hold some money in funds with positive alpha estimates, it is still a concern that this two-step procedure eliminates too many funds that could have been used to enhance diversification gains and increase portfolio performance.

To address this concern, we consider two alternative portfolios. First, we consider a portfolio that includes all funds with positive alpha forecasts. Ignoring estimation error, mean-variance theory suggests that such funds should obtain positive weights in the optimal portfolio (equation (10)). Panel A in Table 5 reports the performance of portfolios comprising all funds with positive alpha estimates. Regardless of whether the funds are weighted by their scaled IR estimates or are assigned equal weights, the risk-adjusted return performance of these portfolios is very low. This suggests that, in practice, alpha estimation error poses an important obstacle towards the successful implementation of mean-variance portfolio theory.

Second, we consider an iterative portfolio approach that directly determines whether funds with small but positive alpha estimates should be included (albeit with smaller weights) in an optimal portfolio due to diversification gains. This iterative fund-weighting approach works as follows. First, we form a portfolio of the n funds in the FCS, using weights that are proportional to their information ratio and, thus, designed to optimize the resulting portfolio's information ratio. Next, we test if this portfolio has a higher IR estimate than that of a portfolio of funds that includes the $n + 1$ ranked fund–the last fund to be eliminated from the FCS–and re-calculate the optimal portfolio weights on this expanded set of funds. If the IR of this augmented portfolio exceeds that of the FCS portfolio by a certain threshold, we include the marginal ($n + 1$) fund and form a portfolio of $n + 1$ funds. Next, we repeat this step, comparing an optimally-weighted portfolio of $n + 1$ funds to an optimally-weighted portfolio of $n + 2$ funds. We proceed until we can reject that the portfolio formed by including $n + k$ funds fails to increase the ex-ante IR value of the portfolio

---

[32]Including the fifth factor in the forecasting and selection steps (Step 1 and Step 3 in Section 2.6) is possible, but not practical. Our performance model already requires estimation of 14 parameters and the inclusion of an additional factor increases this number to 16. This poses a serious challenge to obtaining reliable parameter estimates and would require us to expand the estimation window for the five-factor model.

comprising $n + k - 1$ funds ($k \geq 1$) by the threshold value. We choose the threshold value to be 1% but also consider alternative values of 5% and 10%.

Adding another fund with a non-negative alpha estimate should not, of course, reduce the ex-ante information ratio of the Treynor-Black weighted optimal portfolio. Due to estimation error, however, in practice the ex-post information ratio of the iterative approach may not be higher than that of the (one-shot) FCS approach.

Panel B in Table 5 presents results for this iterative approach using a threshold value of 1%. Empirically, we find that the iterative approach fails to improve on the information ratios of the tight and medium-tight FCS portfolios. However, we now find some (economically modest) improvements in the ex-post information ratios of the medium and wide FCS portfolios constructed using our iterative IR approach compared to the baseline FCS approach. Using higher thresholds such as a 5% or 10% to trigger the addition of funds excluded from the FCS results in portfolios that are very similar to the original FCS portfolios.

To see if estimation error continues to explain these findings, note that the ex-ante expected IR-value for the iterative approach applied to the tight FCS portfolio is 19% higher than the expected IR value for the portfolio based on the original (one-shot) FCS. Hence, we might expect the iterative approach to improve the performance of the FCS-tight portfolio by a sizeable amount (19%). In fact, the actual information ratio of the tight FCS portfolio is lower under the iterative approach as compared to the value for the regular FCS portfolio. This suggests that the additional funds included by the iterative approach have alphas which are estimated with considerable error (and thus do not get included in the FCS set). Such funds may not genuinely have positive alphas even though they are given positive weights by the iterated IR-optimal approach.

These results highlight again that, in practice, estimation error in the information ratios used to form optimal weights makes it difficult to improve on the actual (ex-post) performance of the portfolio constructed from funds included in the FCS.

As discussed previously, institutional investors frequently have restrictions on the number of funds they can invest in. We therefore also consider two portfolio strategies that impose upper or lower bounds on the number of funds, in each case using the ranking information from the sequential FCS approach to determine the identity of the funds held.

For the portfolio restricted to holding a minimum of five funds (Panel C), we find a statistically significant alpha estimate of 33 bps/month for the tight FCS portfolio.[33] The alpha estimate of the portfolio restricted to hold a minimum of 10 funds (20 bps/month; Panel D), is a bit lower, but remains statistically significant in the equal-weighted case and is on the border of being significant (t-statistic of 1.95) with optimal weights. Although investors clearly would have been better off

---

[33]Note that there is a tradeoff between (1) allowing the FCS to completely dictate the elimination, and (2) restricting the FCS to eliminate only up to a certain number of funds, then using a more restrictive ("tight") comparison protocol between remaining funds.

without these restrictions, our FCS approach thus can be used to provide guidance for choosing which funds to include even for investors with constraints on the minimum number of funds to hold.

We next investigate the effect of imposing a cap on the maximum number of funds allowed in the portfolio, $n_{\max}$. At each point in time, if the number of superior funds is smaller than $n_{\max}$, the constrained portfolio consists of all such superior funds. If the number of superior funds exceeds $n_{\max}$, we pick $n_{\max}$ funds at random, repeating this procedure to generate 1,000 portfolios. Panel E reports results in the form of quantiles of the performance distribution of these capped FCS portfolios. Interestingly, the median fund does better when the FCS approach is applied without restricting the number of funds. As shown by the left-tail alpha estimates, the FCS portfolios capped to include one fund are slightly riskier than the wider FCS portfolios capped to hold a maximum of three or five funds. For the FCS portfolios capped at one, three and five funds, respectively, there is a 5% chance of getting an alpha estimate of 48, 57, and 59 bps/month or less. The narrower FCS portfolios also have slightly greater upside, with a 5% chance of obtaining an alpha estimate of at least 85, 78, and 74 bps/month for the FCS portfolios capped at one, three and five funds, respectively. The wider spread of the narrower portfolios is what we would expect due to the effect of diversification across the actively-managed funds. Interestingly, all alpha quantiles–even those in the far left tail (5%)–of these concentrated portfolios remain highly statistically significant. Furthermore, the capped portfolios obtain attractive performance with Sharpe ratios and information ratios of 0.46 or higher even in the far left tail.

In addition to restrictions on the number of funds, some managers are also restricted to certain investment styles. Our approach can also be applied to such cases. For example, applying our approach to choose from large growth funds, the tight FCS set of superior funds generates an alpha estimate of 47 bps/month with a t-statistic of 2.17. Choosing funds from this narrower style set, on average, the superior FCS portfolio includes nine large growth funds.

We conclude from these results that our FCS methodology can be combined with a variety of common constraints on portfolio compositions to produce attractive performance results.

4.8. Fund performance in high and low volatility states

Kacperczyk, van Nieuwerburgh, and Veldkamp (2016) analyze mutual funds' allocation of attention in a setting with multiple risky assets. Assuming that uncertainty about the aggregate risk factors is sufficiently large, their analysis shows that an increase in the variance of the aggregate risk factors increases (i) dispersion in fund portfolio holdings; (ii) dispersion in fund excess returns (Proposition 3); and (iii) excess returns for skilled funds (Proposition 5). Intuitively, rewards from processing information are higher in states where asset payoffs are highly volatile, giving skilled investors stronger incentives to allocate extra attention to their portfolio decisions in such states.

Conversely, unskilled investors can either not increase their attention (noise traders) or face higher costs from doing so, with the result that performance differentials between skilled and unskilled funds increases during periods of high volatility.

To test the first two predictions (wider dispersion in portfolio holdings and returns), we compute the cross-sectional dispersion in industry exposure across the funds in our sample. At each 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.

Panel A in Table 6 (left column) shows that the dispersion in industry concentration–computed using holdings in 48 industries–across funds is marginally higher in high-risk than in low-risk states.[34] However, with a p-value of 0.27, the difference in mean industry concentration across states with high and low aggregate volatility is not statistically significant.

Consistent with the prediction that excess returns should be more dispersed across funds during times of high volatility, the right column of Panel A in Table 6 shows that during periods with high aggregate risk the dispersion in the CS measure is more than three times greater (9.02 versus 2.57) than during periods with low aggregate risk. Moreover, this difference in dispersion across the high and low volatility states is highly 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 due to more dispersion in individual funds' 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 in high-volatility states should enhance our ability to identify superior funds, test procedures are hindered by a larger level of noise in alpha point estimates 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 6 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 (88 bps per month) during periods with high aggregate volatility (left columns) than during periods with low volatility (17 bps/month), and their difference (72 bps/month) is highly statistically significant and more than three times greater than the spread computed using the portfolio of the top 5% alpha-ranked funds ("top 5%"). A similar result holds when we sort by idiosyncratic volatility (right columns) as the alpha of the FCS portfolio is 31

---

[34]A number of 25 corresponds to an average deviation of an industry's weight in a given fund of about 0.7%, relative to the industry's average portfolio weight across funds.

bps/month in low-risk episodes but 75 bps/month during periods with high idiosyncratic risk. The resulting high-low spread (44 bps/month) is more than twice as large as the corresponding spread (17 bps/month) identified using the top 5% of funds.

Panel B2 of Table 6 reports the same results for a 36-month volatility forecast. The results are very similar to those reported for the one-month horizon and the link between volatility and performance is much stronger for the FCS-sorted funds than for portfolios of the top-5% ranked funds.

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 dominates any increased noise affecting the estimated alphas during such periods. The FCS approach effectively identifies the set of superior equity funds during periods when they produce their greatest risk-adjusted return performance. Compared to these findings, the simple alpha-ranked portfolio approach yields notably weaker results in high-volatility states.[35]

5. Selection of superior and inferior funds

Section 4 shows that the FCS approach can be used to endogenously identify the set of funds with superior or inferior risk-adjusted performance. This section provides details of both the number of funds selected by the FCS approach and of their identify.

5.1. Identifying superior funds

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, λ, 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 FCS fluctuates considerably over time. Relatively few funds get selected during the 17-year period from 1993-2010 which coincides with the best risk-adjusted performance of the superior FCS portfolio. At least for this historical period, the FCS methodology thus succeeded in identifying a narrow set of funds with high subsequent performance. During the last 18 months of the sample, the superior FCS set grows in size to include more than 300 funds. This increase is caused by a large number of new funds entering our sample in 2005 and our data requirement of a six-year initial training and estimation sample. On average, across the full sample, the tight FCS ($\lambda = 0.90$) includes 29 funds, while the medium-tight

---

[35]Kacperczyk, van Nieuwerburgh, and Veldkamp (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.

($\lambda = 0.75$), medium ($\lambda = 0.50$) and wide ($\lambda = 0.10$) FCS include 47, 76 and 145 funds, respectively.

Because the number of funds in our data increases over time, the bottom panel in Figure 4 plots the percentage of funds identified by the FCS approach as being superior. For the tight FCS ($\lambda = 0.90$), the percentage of included funds fluctuates between a number close to zero and 20%. For smaller values of λ, the percentage of funds included is, of course, a bit higher. On average, the tight FCS includes 3.5% of the funds, while the medium-tight, medium and wide FCS portfolios include 5.8, 9.5 and 17.0% of the funds, respectively.

For each fund that gets selected at least once during the sample, Figure 5 shows when this particular fund is chosen by the medium tight (top panel) and tight (bottom panel) FCS. (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.) Regardless of the value of λ, we see examples of a single fund being selected for long spells of time. For example, only Fidelity Select Software gets selected during 1993 and 1994 while only T-Rowe Price Media and Telecommunications is selected between 2005 and 2007. For most of the remaining periods a broader set of funds is selected.

We conclude that the set of superior funds identified by our approach fluctuates significantly over time. Sometimes this set is quite broad, containing up to 20% 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.

5.2. Identifying inferior funds

Figure 6 shows that the set of funds with inferior performance fluctuates considerably through time, peaking at close to 30% for the narrow FCS portfolio ($\lambda = 0.90$), at other times containing 1% or fewer of the funds. On average, the tight FCS ($\lambda = 0.90$) includes 89 funds, while the medium-tight ($\lambda = 0.75$), medium ($\lambda = 0.50$) and wide FCS ($\lambda = 0.10$) portfolios on average include 137, 201 and 326 funds, respectively. These figures correspond to 7.6%, 12.9%, 20.1% and 33.4% of the total number of funds, respectively. Thus, across all values of λ, the set of inferior funds is broader and includes more funds than the equivalent FCS for superior funds.

Figure 7 shows inclusion plots for individual funds with inferior performance for the medium tight ($\lambda = 0.75$, top panel) and tight ($\lambda = 0.9$, bottom panel) FCS approach. We continue to see periods during which only a single inferior fund is selected. For example, only the Alger Small Cap Fund gets included by the tight FCS in 1996 and 1997, while, from 1999 and 2000, the tight FCS predominantly consists of the Pacific Advisors Small Cap Fund.

Interestingly, passive index funds, such as the Vanguard S&P500, are never included by either the (tight) superior or inferior FCS portfolios. This suggests that there always exist funds with genuinely better or genuinely worse expected performance than a passive low-cost strategy.

5.3. Market volatility and breadth of the set of superior funds

The model of Kacperczyk, van Nieuwerburgh, and Veldkamp (2016) implies that skilled fund managers should perform better in times of high market volatility. We may therefore expect clearer performance signals and possibly be able to identify a narrower set of superior funds during such times. To see whether the *proportion* of funds identified to have superior performance depends on the state of the economy, we divide the sample into equal-sized terciles according to the level of market volatility.

Panel C of Table 6 shows that more funds are identified as being superior during times of low volatility. Conversely, and consistent with the theoretical predictions of Kacperczyk, van Nieuwerburgh, and Veldkamp (2016), the set of superior funds is narrower in times of high volatility. Moreover, the difference in the proportions of funds selected during high and low volatility states is statistically significant and is robust in the sense that the results hold for both aggregate and idiosyncratic risk measures and regardless of whether we use 12- or 36-months volatility estimates.

5.4. Turnover of funds included in the 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

$$Turnover_t = \frac{1}{2} \sum_{i=1}^{N_t} |\Delta w_{it}|, \tag{21}$$

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 superior FCS portfolio is 0.38, while the corresponding figure for the inferior FCS portfolio is 0.50. For comparison, the monthly turnover of the portfolios of the top 5% of the candidate sets of top and bottom funds are 0.24 and 0.19, respectively. Thus, the turnover is somewhat higher for the FCS portfolios, than for the fractile-based counterparts, which is to be expected as the FCS portfolios consist of more funds on average.

6. Conclusion

This paper presents a new approach to selecting funds with superior performance which 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 pairwise performance comparisons across a large set of mutual

funds, our approach iteratively eliminates funds with inferior performance. 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, consistent with recent studies we find that only a relatively narrow set of funds is capable of outperforming on a risk-adjusted basis. Second, 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 those obtained from simple alpha-ranked portfolios of funds. Clearly there is substantial heterogeneity in performance even among the funds with the highest alpha estimates.

Third, 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.

Fourth, the proportion of funds–as well as the identify of the individual funds–deemed to be superior varies considerably over time. Consistent with predictions from the theoretical analysis in Kacperczyk, van Nieuwerburgh, and Veldkamp (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.

Fifth and finally, the performance of the funds 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.

Internet Appendix A: Performance for different alpha specifications

This appendix shows results for different specifications of the predictive alpha measure in (2). We consider a number of specifications in order to better understand the importance of each component of the predictive alpha specification and their combination. The alternative specifications are listed below with empirical results presented in Table A.1.

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

$$P_{i,t}^{(2)} = R_{i,t} - \hat{\boldsymbol{\beta}}_{i,t}' \mathbf{z}_t$$

$$P_{i,t}^{(3)} = \begin{cases} \hat{\alpha}_{i,t|t-1} \times sign(R_{i,t} - \hat{\boldsymbol{\beta}}_{i,t}' \mathbf{z}_t) & \text{if} \quad -0.02 < \hat{\alpha}_{i,t|t-1} < 0.02 \\ -\infty & \text{otherwise} \end{cases}$$

$$P_{i,t}^{(4)} = \hat{\alpha}_{i,t|t-1} \times sign(R_{i,t} - \hat{\boldsymbol{\beta}}_{i,t}' \mathbf{z}_t) \quad \text{where} \quad \begin{cases} \hat{\alpha}_{i,t|t-1} = q_{\hat{\alpha}}(0.05) & \text{if} \quad q_{\hat{\alpha}}(0.05) > \hat{\alpha}_{i,t|t-1} \\ \hat{\alpha}_{i,t|t-1} = q_{\hat{\alpha}}(0.95) & \text{if} \quad q_{\hat{\alpha}}(0.95) < \hat{\alpha}_{i,t|t-1} \end{cases}$$

$$P_{i,t}^{(5)} = (R_{i,t} - \hat{\boldsymbol{\beta}}_{i,t}' \mathbf{z}_t) \times \left(\frac{2}{1 + exp(-\gamma \hat{\alpha}_{i,t|t-1})} - 1 \right)$$

$$P_{i,t}^{(6)} = (R_{i,t} - \hat{\boldsymbol{\beta}}_{i,t}' \mathbf{z}_t) \times sign(\hat{\alpha}_{i,t|t-1}) \times exp(-\gamma|\hat{\alpha}_{i,t|t-1}|)$$

$$P_{i,t}^{(7)} = (R_{i,t} - \hat{\boldsymbol{\beta}}_{i,t}' \mathbf{z}_t) \times \left(2\Phi(\hat{\alpha}_{i,t|t-1}) - 1 \right)$$

$$P_{i,t}^{(8)} = (R_{i,t} - \hat{\boldsymbol{\beta}}_{i,t}' \mathbf{z}_t)\hat{\alpha}_{i,t|t-1} \quad \text{where} \quad \begin{cases} \hat{\alpha}_{i,t|t-1} = q_{\hat{\alpha}}(0.05) & \text{if} \quad q_{\hat{\alpha}}(0.05) > \hat{\alpha}_{i,t|t-1} \\ \hat{\alpha}_{i,t|t-1} = q_{\hat{\alpha}}(0.95) & \text{if} \quad q_{\hat{\alpha}}(0.95) < \hat{\alpha}_{i,t|t-1} \end{cases}$$

Here $q_{\hat{\alpha}}(p)$ denotes the p-th quantile of the cross-sectional distribution of $\hat{\alpha}$, $\Phi(x)$ is the CDF of a standard normal distribution, and γ is a tuning parameter.

The first specification only uses the alpha forecast. This model performs poorly and leads to a very small number of funds being selected with estimation noise and outliers dominating the fund selection process. The second specification simply uses the risk-adjusted fund return, controlling for common risk factors. Again, this approach performs poorly because there is no signal about manager skill. The third and fourth specifications are based on the alpha forecast multiplied by the sign of the realized return and thus represent different ways of dealing with outliers. The third model eliminates funds with large forecasts, using the elimination rule introduced in Mamaysky, Spiegel, and Zhang (2007). This improves results considerably, but the performance remains insignificant in statistical terms. The fourth model uses a winsorization strategy which produces significant performance for the tight and medium-tight specifications. These results indicate that it is important to limit the influence of outliers in the alpha estimates. Our predictive alpha specification in (2) accomplishes this by utilizing only the sign of the forecast, thus clearly limiting

the influence of outliers in alpha forecasts since all positive (negative) forecasts receive the same weight. Models 5-8 present alternatives ways to reduce the sensitivity to extreme alpha estimates. Model 5 places smaller weights on the returns of funds with low alpha forecasts while also ensuring that funds with implausibly high alpha forecasts do not receive too large weights. Model 6 puts lower weights on very high forecasts of alpha. Model 7 uses the CDF of the normal distribution to assign weights, ensuring that small alpha forecasts receive a low weight while, conversely, large alpha forecasts are not weighted too heavily. Finally, Model 8 uses a winsorization scheme in place of the sign function used in our specification in (2). Specifications 5-8 perform very similarly to our preferred specification. Some perform better and some perform worse. We conclude from this that the performance documented in the main part of the manuscript is not unique to the specification in (2), but reflects a broader finding about the importance of exploiting predictability in risk-adjusted returns while simultaneously controlling for outliers in the alpha forecasts.

As illustrated in this appendix, it is important to limit the influence of the largest alpha estimates. This can also be true when forming portfolios based on funds with extreme IRs which tend to receive very large weights. A simple way to deal with this issue is to winsorize the distribution of IRs included in the FCS to limit the influence of the funds with the largest ratios. This winsorization scheme is only used if explicitly stated by labelling the results as "optimally weighted portfolio (winsorized)."

Internet Appendix B: Alternative fund selection algorithm

This appendix considers the properties of the MCS approach–which requires conducting an all possible pairwise model comparisons–and the simpler "top fund" approach which identifies the best-performing model and eliminates as many models as it can by comparing them against this fund.

Since analytical results are generally not feasible, we use Monte Carlo simulations to compare the two approaches. Our simulation study uses the following setup. Consider 100 funds whose mean returns are chosen such that N_s of the funds are superior in the sense that they have higher means. We accomplish this by letting the first N_s funds have a mean return of 1 while the remaining $100 - N_s$ funds have a mean return of 0.25. All fund returns are normally distributed with a standard deviation of 1 and are uncorrelated across funds. Our analysis assumes a 10% significance level (size).

Our simulations check whether the first N_s funds—those whose mean returns are genuinely higher—are selected as being superior by the MCS methodology. For the "top comparison" procedure, we first identify the fund with the highest mean return, then calculate all 99 t–statistics by comparing the top fund to the remaining ones. Finally, we eliminate all funds with a t–statistic significantly greater than the critical value of a one-sided t–test; since we use the highest mean return, we know that all t–statistics are positive and so can apply a one-sided test. Next, we

consider the fund with the second highest mean and repeat the elimination steps. We continue until no more funds are eliminated. Finally, we check whether the first N_s funds remain in the set of funds that have not been eliminated. We repeat this procedure 100 times to construct 100 sets of selected funds for the two methods.

Figure B.1 shows the fraction of the 100 simulations in which all the truly superior funds (N_s) are included as we vary the number of truly superior funds (shown on the x-axis) from $N_s = 1$ to $N_s = 10$. The top and bottom panels assume sample sizes of 60 and 120 observations, respectively. The red line tracks the level of the FCS; the blue line tracks the level of the iterative elimination procedure. The green line is simply one minus the nominal level of 10% (0.90) and serves as a reference point as tests with the right size should closely mirror this line.

We observe the following. In the presence of a single superior fund ($N_s = 1$), both the FCS and the "top comparison" methods perform very well. However, as the number of truly superior funds (N_s) increases, the performance of the iterative elimination approach deteriorates in the sense that it fails to correctly include all N_s funds. Conversely, the FCS approach continues to perform well regardless of whether N_s is small or large.

Internet Appendix C: Estimation and prediction of fund performance

This appendix explains how we estimate the unknown parameters of the latent skill 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:

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

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 ith 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:

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

Using information up to time $t-1$, we can estimate the parameters of the latent skill 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, $i = 1, ..., N_t$, 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 D: 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:

$$
R_{pt} = \alpha_p(t/T) + \boldsymbol{\beta}'_p(t/T)\mathbf{z}_t + \varepsilon_{pt}, \quad t = 1, \ldots, T. \tag{22}
$$

Here $\alpha_p(\cdot)$ and $\boldsymbol{\beta}_p(\cdot)$ are unknown smooth functions that are allowed to depend on the sample "fraction", t/T, and, thus, can vary over time.[36] To see how we can nonparametrically estimate the parameters, define the vector of regressors $\mathbf{X}_t = (1\ \mathbf{z}_t)$ and parameters $\boldsymbol{\theta}_p(t/T) = (\alpha_p(t/T)$

---

[36]Technically, the $\alpha_p(\cdot)$ and $\boldsymbol{\beta}_p(\cdot)$ functions allow for a finite number of discontinuities.

$\boldsymbol{\beta}'_p(t/T))'$, and rewrite equation (22) as

$$R_{pt} = \mathbf{X}'_t \boldsymbol{\theta}_p(t/T) + \varepsilon_{pt}. \tag{23}$$

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

$$k_{st}^{1/2} R_{ps} = k_{st}^{1/2} \mathbf{X}'_s \boldsymbol{\gamma}_{p0} + k_{st}^{1/2} \left(\frac{s-t}{T} \right) \mathbf{X}'_s \boldsymbol{\gamma}_{p1} + \varepsilon_{ps}, \quad s = 1, \ldots, T, \tag{24}$$

where $k_{st} = k(\frac{s-t}{Th})$ is a kernel function, and h is the bandwidth.[37] In the second step, we construct an estimator of $\boldsymbol{\alpha}_{pt} = \boldsymbol{\alpha}_p(t/T)$ as

$$\hat{\boldsymbol{\alpha}}_{pt} = (\mathbf{e} \otimes \mathbf{I})\hat{\boldsymbol{\gamma}}_p, \tag{25}$$

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.

The estimate $\hat{\alpha}_{p,t}$ from (25) portrays the evolution in the performance of the different portfolios in a way that does not "average out" potentially interesting time-variation in α_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.

---

[37]Following Chen and Hong (2012), we set $h = T^{-1/5}/\sqrt{12}$.

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

Berk, J.B., Green, R.C., 2004, Mutual fund flows and performance in rational markets. Journal of Political Economy 112, 1269-1295.

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. The 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., Petajisto, A., 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. The 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.

Fama, E.F., French, K.R., 1993, Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 427–465

Ferson, W.E., Schadt, R.W., 1996, Measuring fund strategy and performance in changing economic conditions. The Journal of Finance 51, 425–461.

Garleanu, N., Pedersen, L.H., 2018, Efficiently inefficient markets for assets and asset management. The Journal of Finance 73, 1663–1712.

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., Liu, Y., 2018, False (and Missed) Discoveries in financial economics. Unpublished working paper, Duke.

Harvey, C.R., Liu, Y., 2019, Cross-sectional alpha dispersion and performance evaluation. Journal of Financial Economics, Forthcoming.

Hoberg, G., Kumar, N., Prabhala, N., 2018. Mutual fund competition, managerial skill, and alpha persistence, Review of Financial Studies 31, 1896–1929.

Kacperczyk, M., Sialm, C., Zheng, L., 2005, On the industry concentration of actively managed equity mutual funds. The Journal of Finance 60, 1983-2012.

Kacperczyk, M., Sialm, C., Zheng, L., 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. The Journal of Finance 69, 1455-1484.

Kacperczyk, M., van Nieuwerburgh, S., Veldkamp, L., 2016, A rational theory of mutual funds' attention allocation. Econometrica 84(2), 571-626.

Kosowski, R., Timmermann, A., Wermers, R., and White, H., 2006, Can Mutual Fund Stars Really Pick Stocks? New Evidence from a Bootstrap Analysis. The Journal of Finance 61, 2551–2595.

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.

Pastor, L., Stambaugh, R.F., Taylor, L.A., 2015, Scale and skill in active management. Journal of Financial Economics 116, 23-45.

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

Wei, K., Wermers, R., Yao, T., 2015, Uncommon value: The investment performance of contrarian funds. Management Science, October, 2394–2414.

Wermers, R., 1999, Mutual Fund Herding and the Impact on Stock Prices. The Journal of Finance 54, 581–622.

Wermers, R., 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. The 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

Panel A: alpha (x100)

	Shearson App.	Growth F. of Am.	Sogen Int. F.	Pioneer II
α	0.14	0.46	0.54	0.10
Superior	✓	✓	✓	✓

Panel B: t-statistics of alpha

	Shearson App.	Growth F. of Am.	Sogen Int. F.	Pioneer II
t_α	2.11	2.98	2.45	0.70
Superior	✓	✓	✓	−

Panel C: Pair-wise t-statistics (FCS methodology)

Step 1 - All

	Shearson App.	Growth F. of Am.	Sogen Int. F.	Pioneer II
Shearson App.	0			
Growth F. of Am.	2.03	0		
Sogen Int. F.	1.67	0.29	0	
Pioneer II.	-0.28	-1.81	-2.04	0

Step 2 - Pioneer II removed

	Shearson App.	Growth F. of Am.	Sogen Int. F.
Shearson App.	0		
Growth F. of Am.	2.03	0	
Sogen Int. F.	1.67	0.29	0

Step 3 - Pioneer II and Shearson App.

	Growth F. of Am.	Sogen Int. F.
Growth F. of Am.	0	
Sogen Int. F.	0.29	0

	Shearson App.	Growth F. of Am.	Sogen Int. F.	Pioneer II
Superior	−	✓	✓	−

The table presents an example of how different fund selection methodologies identify superior funds applied to actual data for four funds obtained in November 1990. 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		
	alpha	SR	IR
Mean	-0.072	0.309	-0.349
5%	-0.436	-0.051	-1.919
10%	-0.322	0.043	-1.378
15%	-0.180	0.200	-0.804
50%	-0.065	0.333	-0.309
75%	0.037	0.439	0.166
90%	0.157	0.535	0.646
95%	0.258	0.592	1.006

This table shows the cross-sectional distribution of four-factor alpha estimates (in percent per month), sharpe ratio (annualized), and information ratio (annualized) obtained from monthly returns data on U.S. equity mutual funds over the period 1986:07-2012:12. 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. The cross-sectional distribution has been winsorized at 1%.

Table 3: Risk-adjusted performance for alpha-ranked and predictive alpha-ranked portfolios

	alpha	t-stat	SR	IR	Av. no. funds	alpha	t-stat	SR	IR	Av. no. funds
					Latent skill holdings					
			Alpha sorted					Predictive alpha sorted		
Optimally weighted portfolios										
Pos. Alpha	-0.01	-0.50	0.38	-0.11	351	-0.01	-0.50	0.38	-0.11	351
Pos. Pred Alpha	0.00	0.09	0.40	0.02	183	0.00	0.09	0.40	0.02	183
Top 10%	-0.05	-0.54	0.35	-0.13	19	0.12	1.20	0.43	0.27	19
Top 5%	0.02	0.20	0.40	0.04	10	0.25	1.95	0.48	0.44	10
Equal weighted portfolios										
Pos. Alpha	0.01	0.13	0.40	0.03	351	0.01	0.13	0.40	0.03	351
Pos. Pred Alpha	0.04	0.83	0.43	0.19	183	0.04	0.83	0.43	0.19	183
Top 10%	0.14	1.56	0.45	0.36	19	0.15	1.56	0.43	0.35	19
Top 5%	0.17	1.56	0.47	0.34	10	0.29	2.25	0.47	0.50	10
					Latent skill					
			Alpha sorted					Predictive alpha sorted		
Optimally weighted portfolios										
Pos. Alpha	-0.02	-0.15	0.33	-0.03	404	-0.02	-0.15	0.33	-0.03	404
Pos. Pred Alpha	-0.12	-1.01	0.25	-0.19	209	-0.12	-1.01	0.25	-0.19	209
Top 10%	-0.03	-0.28	0.34	-0.06	21	0.20	1.51	0.49	0.31	21
Top 5%	0.02	0.18	0.37	0.04	11	0.33	2.13	0.53	0.45	11
Equal weighted portfolios										
Pos. Alpha	-0.00	-0.09	0.39	-0.02	404	-0.00	-0.09	0.39	-0.02	404
Pos. Pred Alpha	0.04	0.98	0.42	0.22	209	0.04	0.98	0.42	0.22	209
Top 10%	0.14	1.60	0.44	0.35	21	0.16	1.62	0.42	0.35	21
Top 5%	0.06	0.69	0.42	0.15	11	0.26	1.88	0.43	0.42	11
					Constant skill					
			Alpha sorted					Predictive alpha sorted		
Optimally weighted portfolios										
Pos. Alpha	0.01	0.39	0.40	0.09	415	0.01	0.39	0.40	0.09	415
Pos. Pred Alpha	0.02	0.58	0.41	0.12	223	0.02	0.58	0.41	0.12	223
Top 10%	0.16	1.81	0.40	0.39	23	0.19	1.88	0.44	0.43	23
Top 5%	0.30	2.12	0.42	0.48	12	0.25	1.90	0.44	0.44	12
Equal weighted portfolios										
Pos. Alpha	0.01	0.26	0.40	0.06	415	0.01	0.26	0.40	0.06	415
Pos. Pred Alpha	0.02	0.51	0.41	0.11	223	0.02	0.51	0.41	0.11	223
Top 10%	0.22	1.90	0.40	0.42	23	0.22	1.98	0.44	0.45	23
Top 5%	0.35	2.19	0.42	0.50	12	0.26	1.91	0.44	0.44	12

This table reports four-factor alphas (in percent per month) for a set of alpha-ranked and predictive alpha-ranked portfolios. Each month we rank the set of mutual funds with positive expected alpha and positive average predictive alpha according to their expected alpha for the next month ahead and according to their average predictive alphas. We consider two different sets of portfolio weights. One, where the weight on a fund i is determined by $\frac{\hat{\alpha}_{i,t+h|t}/\hat{\sigma}_i^2}{\sum_{j=1}^n \hat{\alpha}_{j,t+h|t}/\hat{\sigma}_j^2}$, for $h=1$ and n funds, which is labeled optimally weighted and one, where funds are equal weighted. We present results for the portfolio of all funds with positive forecasts of alpha (labelled Pos. Alpha), all funds with positive forecasts and positive average predictive alpha (the candidate set, labelled Pos. Pred. Alpha), and narrow subsets of the candidate set. We present results for a portfolio of the 10% highest ranking funds in the candidate set, as well as the top 5% highest ranking funds. The ranking is repeated each month during the sample 1986:07-2012:12 and so produces a time series of portfolio returns from which the reported alpha estimates are computed. The table reports results for the model with constant skill, latent skill, and the latent skill holdings model, which allows for a time-varying alpha and combines returns and holdings data to estimate each fund's alpha. Results for the alpha-sorted portfolios are presented on the left hand side of the table, while results for the predictive alpha-sorted portfolios are on the right hand side of the table. The table reports monthly alpha estimates in percentage terms followed by t-statistics, sharpe ratio (annualized), information ratio (annualized), and the average number of funds included in the portfolio.

Table 4: Risk-adjusted performance for FCS portfolios

Percent in top	alpha	t-stat	SR	IR	Av. no. funds	alpha	t-stat	SR	IR	Av. no. funds
					Latent skill holdings					
			Superior funds					Inferior funds		
Optimally weighted portfolios										
FCS-tight	0.66	3.25	0.61	0.74	29	-0.34	-2.51	0.16	-0.62	89
FCS-medium-tight	0.31	1.92	0.49	0.43	47	-0.21	-2.38	0.23	-0.46	137
FCS-medium	0.06	0.59	0.40	0.12	76	-0.16	-2.10	0.29	-0.42	201
FCS-wide	-0.01	-0.21	0.39	-0.05	145	-0.15	-3.61	0.31	-0.81	326
Equal weighted portfolios										
FCS-tight	0.67	3.41	0.60	0.77	29	-0.31	-2.40	0.17	-0.58	89
FCS-medium-tight	0.33	2.07	0.48	0.46	47	-0.19	-2.31	0.24	-0.43	137
FCS-medium	0.10	0.97	0.41	0.21	76	-0.12	-1.70	0.31	-0.33	201
FCS-wide	0.04	0.78	0.43	0.18	145	-0.12	-2.88	0.32	-0.64	326
Optimally weighted portfolios (Winsorized)										
FCS-tight	0.67	3.32	0.62	0.76	29	-0.34	-2.51	0.16	-0.62	89
FCS-medium-tight	0.33	2.02	0.50	0.45	47	-0.21	-2.37	0.23	-0.46	137
FCS-medium	0.08	0.84	0.42	0.18	76	-0.16	-2.09	0.29	-0.42	201
FCS-wide	0.02	0.29	0.42	0.07	145	-0.15	-3.57	0.31	-0.80	326
					Latent skill					
			Superior funds					Inferior funds		
Optimally weighted portfolios										
FCS-tight	-0.08	-0.43	0.34	-0.09	54	-0.38	-2.03	0.11	-0.47	104
FCS-medium-tight	-0.04	-0.23	0.34	-0.05	78	-0.38	-2.48	0.13	-0.54	150
FCS-medium	-0.04	-0.27	0.30	-0.05	110	-0.28	-1.94	0.17	-0.42	210
FCS-wide	-0.11	-0.88	0.26	-0.17	171	-0.13	-1.22	0.28	-0.26	323
Equal weighted portfolios										
FCS-tight	0.20	1.32	0.47	0.29	54	-0.42	-2.52	0.10	-0.63	104
FCS-medium-tight	0.17	1.42	0.44	0.32	78	-0.34	-2.61	0.16	-0.61	150
FCS-medium	0.06	0.72	0.40	0.15	110	-0.25	-2.22	0.20	-0.51	210
FCS-wide	0.02	0.43	0.40	0.09	171	-0.12	-2.77	0.32	-0.63	323
Optimally weighted portfolios (Winsorized)										
FCS-tight	0.15	1.01	0.46	0.23	54	-0.38	-2.02	0.11	-0.47	104
FCS-medium-tight	0.14	1.15	0.45	0.26	78	-0.38	-2.46	0.13	-0.54	150
FCS-medium	0.05	0.68	0.42	0.14	110	-0.28	-1.95	0.17	-0.42	210
FCS-wide	0.02	0.31	0.42	0.07	171	-0.13	-1.22	0.28	-0.26	323
					Constant skill					
			Superior funds					Inferior funds		
Optimally weighted portfolios										
FCS-tight	0.32	1.88	0.49	0.40	42	-0.23	-2.90	0.24	-0.53	118
FCS-medium-tight	0.28	1.97	0.49	0.41	64	-0.21	-3.52	0.25	-0.75	163
FCS-medium	0.09	0.98	0.42	0.20	98	-0.20	-3.96	0.26	-0.89	218
FCS-wide	0.03	0.64	0.42	0.14	180	-0.15	-3.74	0.31	-0.89	327
Equal weighted portfolios										
FCS-tight	0.34	1.97	0.50	0.42	42	-0.21	-2.49	0.25	-0.47	118
FCS-medium-tight	0.27	1.83	0.47	0.40	64	-0.18	-2.99	0.27	-0.65	163
FCS-medium	0.10	1.02	0.41	0.21	98	-0.17	-3.27	0.28	-0.76	218
FCS-wide	0.03	0.50	0.41	0.11	180	-0.13	-2.97	0.32	-0.73	327
Optimally weighted portfolios (Winsorized)										
FCS-tight	0.32	1.87	0.49	0.40	42	-0.23	-2.89	0.24	-0.53	118
FCS-medium-tight	0.28	1.95	0.48	0.41	64	-0.21	-3.50	0.25	-0.75	163
FCS-medium	0.09	1.03	0.42	0.21	98	-0.20	-3.93	0.27	-0.88	218
FCS-wide	0.03	0.52	0.42	0.11	180	-0.15	-3.72	0.31	-0.89	327

Caption on next page

Caption for Table 4. This table reports four-factor alphas for a set of portfolios of funds identified by the fund confidence set (FCS) approach. Alpha estimates are in percent per month and presented with corresponding t-statistics. The table also presents sharpe ratio (annual), information ratio (annual), and the average number of funds included in the portfolios. Each month we apply the FCS approach to all funds with positive (negative) predictive alpha one month ahead to identify the set of superior (inferior) funds. We then form a portfolio of these funds and record its return during the following month. We consider three different sets of portfolio weights. One where the portfolio weight on fund i is given by $\frac{\hat{\alpha}_{i,t+1|t}/\hat{\sigma}_i^2}{\sum_{j=1}^{n}\hat{\alpha}_{j,t+1|t}/\hat{\sigma}_j^2}$, which is labeled optimally weighted, one, where funds are equal weighted, and one, where the portfolio is optimally weighted, but the weights are winsorized to limit the influence of the 10 percent highest information ratios. This procedure is repeated each month during the sample 1986:07-2012:12 and so produces a time series of portfolio returns from which the reported four-factor alpha estimates are computed. The left panel presents results for the set of superior funds, while the right panel presents results for the set of inferior funds. In each panel the four rows report results for four different values of the tightness parameter (λ) 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. The table presents results for a model assuming constant skill, the latent skill model, and 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: Risk-adjusted performance for alternative portfolios

	alpha	t-stat	SR	IR	Av. no. funds	alpha	t-stat	SR	IR	Av. no. funds
Panel A: Funds with positive forecasts										
			Optimally weighted					Equal weighted		
Positive forecast	-0.01	-0.50	0.38	-0.11	351	0.01	0.13	0.40	0.03	351
Panel B: FCS with iterative portfolio approach										
			Optimally weighted					Equal weighted		
FCS-tight	0.44	3.04	0.55	0.67	31	0.50	3.58	0.56	0.79	31
FCS-medium-tight	0.22	1.88	0.47	0.39	48	0.23	2.00	0.45	0.44	48
FCS-medium	0.07	0.93	0.42	0.19	77	0.10	1.20	0.42	0.25	77
FCS-wide	-0.01	-0.11	0.39	-0.02	145	0.04	0.79	0.43	0.19	145
Panel C: FCS with minimum five funds										
			Optimally weighted					Equal weighted		
FCS-tight	0.33	2.48	0.50	0.54	31	0.31	2.58	0.47	0.56	31
FCS-medium-tight	0.12	1.02	0.41	0.22	48	0.15	1.38	0.41	0.30	48
FCS-medium	0.05	0.72	0.41	0.14	76	0.07	0.95	0.40	0.19	76
FCS-wide	-0.01	-0.21	0.39	-0.05	145	0.04	0.78	0.43	0.18	145
Panel D: FCS with minimum ten funds										
			Optimally weighted					Equal weighted		
FCS-tight	0.20	1.95	0.43	0.41	33	0.20	2.01	0.42	0.44	33
FCS-medium-tight	0.09	1.03	0.39	0.21	49	0.11	1.27	0.38	0.27	49
FCS-medium	0.03	0.38	0.39	0.08	77	0.06	0.86	0.40	0.18	77
FCS-wide	-0.01	-0.21	0.39	-0.05	145	0.04	0.78	0.43	0.18	145
Panel E: Quantiles of performance distribution for FCS with cap on maximum number of funds										
			Optimally weighted					Equal weighted		
			Maximum one fund					Maximum one fund		
5%	0.48	2.11	0.47	0.46		0.48	2.11	0.47	0.46	
10%	0.52	2.24	0.49	0.49		0.52	2.24	0.49	0.49	
25%	0.59	2.55	0.53	0.56		0.59	2.55	0.53	0.56	
50%	0.67	2.93	0.57	0.64		0.67	2.93	0.57	0.64	
75%	0.75	3.28	0.61	0.71		0.75	3.28	0.61	0.71	
90%	0.81	3.59	0.65	0.78		0.81	3.59	0.65	0.78	
95%	0.85	3.78	0.67	0.82		0.85	3.78	0.67	0.82	
			Maximum three funds					Maximum three funds		
5%	0.57	2.69	0.55	0.60		0.59	2.84	0.54	0.63	
10%	0.59	2.81	0.56	0.63		0.61	2.95	0.55	0.66	
25%	0.63	2.97	0.58	0.67		0.64	3.10	0.57	0.69	
50%	0.67	3.19	0.61	0.72		0.67	3.30	0.59	0.74	
75%	0.71	3.39	0.63	0.76		0.71	3.48	0.61	0.78	
90%	0.75	3.59	0.65	0.81		0.75	3.64	0.63	0.81	
95%	0.78	3.72	0.66	0.83		0.77	3.77	0.64	0.84	
			Maximum five funds					Maximum five funds		
5%	0.59	2.83	0.57	0.64		0.61	2.99	0.56	0.68	
10%	0.61	2.91	0.58	0.66		0.62	3.09	0.57	0.70	
25%	0.63	3.05	0.59	0.69		0.65	3.21	0.58	0.72	
50%	0.66	3.21	0.61	0.73		0.67	3.35	0.59	0.75	
75%	0.70	3.37	0.62	0.76		0.70	3.49	0.61	0.78	
90%	0.72	3.50	0.64	0.79		0.73	3.62	0.62	0.81	
95%	0.74	3.60	0.65	0.81		0.74	3.69	0.63	0.83	

Caption on next page

Caption for Table 5. The table presents risk-adjusted performance for different portfolios of funds. Panel A presents results for a portfolio of funds chosen each month to consist of all funds with a positive alpha forecast for the next month. For each month we estimate the FCS of superior funds and construct a portfolio of the funds identified to be superior. Panel B presents results for a iterative approach, where, at each month, we include the last fund to be eliminated in the portfolio and adjust the portfolio weights accordingly. If the extra fund improves the information ratio of the portfolio more than a pre-specified threshold we keep the fund in the portfolio and include yet another fund. We continue until the information ratio of the portfolio does not increase more than the threshold. Results are tabulated for a threshold of one percent. In panel C and D we present results for a constrained version of the FCS where the methodology each month selects a minimum of five (Panel C) or ten (Panel D) funds respectively. Panel E presents results from imposing a cap on the maximum number of funds included. When too many funds are included in the FCS we chose funds from the set at random until we reach the cap. We construct 1.000 of such portfolios and this panel presents selected quantiles of the performance distributions from imposing different caps.

Table 6: Dispersion and risk-adjusted performance sorted by volatility

Panel A: Dispersion in Industry Concentration and CS Measure

	Industry concentration	CS measure
	Aggregate risk	Aggregate risk
	All funds	All funds
Low risk	24.943 [0.000]	2.573 [0.000]
Med. risk	25.202 [0.000]	4.669 [0.000]
High risk	25.168 [0.000]	9.022 [0.000]
High-Low	0.226 [0.268]	6.449 [0.000]

Panel B1: Dispersion of performance for superior funds (12 months)

	Aggregate risk		Idiosyncratic risk	
	FCS	Top 5%	FCS	Top 5%
Low risk	0.165 [0.031]	-0.125 [0.004]	0.307 [0.001]	-0.074 [0.121]
Med. risk	0.261 [0.000]	0.017 [0.736]	0.251 [0.000]	-0.093 [0.047]
High risk	0.880 [0.000]	0.035 [0.564]	0.749 [0.000]	0.094 [0.122]
High-Low	0.715 [0.000]	0.160 [0.032]	0.442 [0.002]	0.168 [0.030]

Panel B2: Dispersion of performance for superior funds (36 months)

	Aggregate risk		Idiosyncratic risk	
	FCS	Top 5%	FCS	Top 5%
Low risk	0.093 [0.004]	-0.077 [0.000]	0.145 [0.001]	-0.002 [0.930]
Med. risk	0.363 [0.000]	0.018 [0.500]	0.244 [0.000]	-0.095 [0.000]
High risk	1.047 [0.000]	0.236 [0.000]	1.115 [0.000]	0.273 [0.000]
High-Low	0.954 [0.000]	0.313 [0.000]	0.970 [0.000]	0.275 [0.000]

Panel C: Breadth of fund confidence set of superior funds across volatility states

	Aggregate risk		Idiosyncratic risk	
	12 months	36 months	12 months	36 months
All	0.035	0.035	0.035	0.035
Low risk	0.035	0.041	0.042	0.050
Med. risk	0.049	0.046	0.040	0.044
High risk.	0.023	0.024	0.025	0.016
High-Low	-0.013	-0.017	-0.018	-0.034
p-value	[0.062]	[0.025]	[0.016]	[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 β_t is the average market beta across funds and σ_{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 36 monthly observations. Dispersion in industry concentration is calculated as $\frac{1}{48}\sum_{c=1}^{48}(w_{i,t}^c - w_{avg,t}^c)^2$, 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 and B2 report average alpha performance of portfolios of superior funds identified by the fund confidence set (FCS) and the top five percent funds from the benchmark sorting approach sorted according to the rolling 12-month and 36-month volatility estimates, respectively. Average alpha estimates are reported in percent per month and corresponding p-values are presented in brackets. Panel C presents means of the percentage of funds identified by the tight FCS approach as being superior across different volatility states. The percentage of funds selected are presented for all periods, for periods with high volatility only, periods for medium levels of volatility only, and for periods with low volatility only. The row labelled High-Low presents the difference between the breath of the set in high and low volatility period and the row labelled p-value refers to the test that equal percentages of funds are included in high and low volatility periods. Each month we apply the FCS approach to all funds with postive expected alpha one and three months ahead and a positive average predictive alpha in order to identify the set of superior funds. We then form a portfolio of these funds, where the weight on fund i is given by $\frac{\hat{\alpha}_{i,t+1|t}/\hat{\sigma}_i^2}{\sum_{j=1}^{n}\hat{\alpha}_{j,t+1|t}/\hat{\sigma}_j^2}$, and record its return during the following month. This procedure is repeated each month during the sample 1986:07-2012:12 and so produces a time series of portfolio returns from which four-factor alpha estimates are computed based on the same 12- or 36-month rolling window used to calculate the volatility measures.

Table A.1: Alternative specifications of predictive alpha

	alpha	t-stat	SR	IR	Av. no. funds	alpha	t-stat	SR	IR	Av. no. funds
Panel A: $P_{i,t} = P_{i,t}^{(1)}$						**Panel B**: $P_{i,t} = P_{i,t}^{(2)}$				
FCS-tight	-0.12	-0.97	0.25	-0.19	1	0.17	1.13	0.37	0.24	59
FCS-medium-tight	-0.13	-1.09	0.26	-0.21	2	0.12	1.00	0.39	0.22	80
FCS-medium	-0.13	-1.19	0.24	-0.23	2	0.09	0.97	0.40	0.22	105
FCS-wide	-0.09	-0.85	0.28	-0.17	3	0.03	0.53	0.40	0.12	160
Panel C: $P_{i,t} = P_{i,t}^{(3)}$						**Panel D**: $P_{i,t} = P_{i,t}^{(4)}$				
FCS-tight	0.24	1.55	0.47	0.34	18	0.30	2.32	0.56	0.53	26
FCS-medium-tight	0.15	1.15	0.45	0.27	34	0.23	2.45	0.54	0.53	42
FCS-medium	0.11	1.12	0.45	0.25	55	0.09	1.30	0.46	0.28	65
FCS-wide	0.04	0.56	0.42	0.14	115	0.04	0.81	0.43	0.19	126
Panel E: $P_{i,t} = P_{i,t}^{(5)}$						**Panel F**: $P_{i,t} = P_{i,t}^{(6)}$				
$\gamma = 10$						$\gamma = 2$				
FCS-tight	0.59	2.86	0.56	0.65	26	0.71	3.58	0.61	0.81	29
FCS-medium-tight	0.37	2.26	0.49	0.51	43	0.35	2.13	0.48	0.48	47
FCS-medium	0.22	1.69	0.46	0.36	70	0.07	0.72	0.39	0.15	77
FCS-wide	0.04	0.64	0.43	0.15	141	0.04	0.76	0.43	0.18	145
$\gamma = 5$						$\gamma = 1$				
FCS-tight	0.65	3.26	0.61	0.73	23	0.70	3.53	0.61	0.80	29
FCS-medium-tight	0.40	2.35	0.51	0.53	40	0.34	2.17	0.48	0.48	47
FCS-medium	0.16	1.40	0.43	0.29	67	0.06	0.63	0.39	0.14	76
FCS-wide	0.04	0.74	0.43	0.18	137	0.04	0.78	0.43	0.18	145
$\gamma = 1$						$\gamma = 0.5$				
FCS-tight	0.51	2.65	0.52	0.59	15	0.68	3.47	0.60	0.79	29
FCS-medium-tight	0.54	3.24	0.60	0.71	29	0.33	2.07	0.48	0.46	47
FCS-medium	0.21	1.91	0.46	0.41	53	0.10	0.96	0.41	0.21	76
FCS-wide	0.04	0.54	0.42	0.13	113	0.04	0.78	0.43	0.19	145
Panel G: $P_{i,t} = P_{i,t}^{(7)}$						**Panel H**: $P_{i,t} = P_{i,t}^{(8)}$				
FCS-tight	0.60	2.94	0.55	0.66	27	0.47	2.40	0.53	0.53	14
FCS-medium-tight	0.39	2.38	0.50	0.54	44	0.51	3.20	0.59	0.73	28
FCS-medium	0.25	1.97	0.47	0.42	72	0.22	1.99	0.50	0.42	52
FCS-wide	0.04	0.74	0.43	0.17	142	0.06	0.86	0.44	0.20	111

The table presents risk-adjusted performance for equal weighted portfolios based on eight alternative specifications of the predictive alpha. Panel A presents results for a specification, where the predictive alpha is simply the forecast of alpha. In Panel B, the predictive alpha is specified as the realized return for a fund, controlled for common risk factors. Panel C and Panel D present results for a specification, where the forecast of alpha is multiplied by 1 or -1 depending on, whether the realized return is positive or negative, respectively. The two panel deal with outliers in different ways. Panel E-H present different alternatives to our main specifiction presented in (2). Each month we estimate the FCS based on a particular specification of predictive alpha. We form a portfolio of the funds identified as superior and track the performance over the following month.
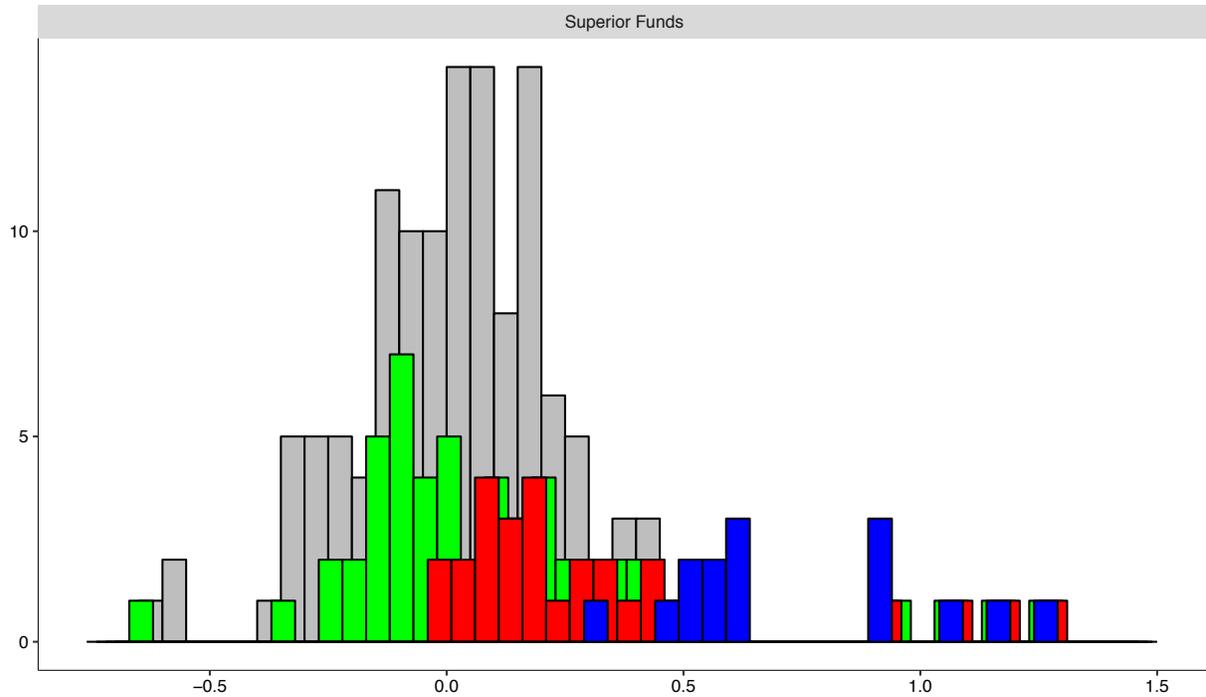
Figure 1: Predictive alpha for top-ranked funds. The figure presents distributions of predictive alphas for funds included in various portfolios at a single month in our sample (November 1994). The figure presents distributions of average predicted alphas for the top-ranked funds. The gray bars represents a portfolio consisting of all funds in existence in December 1994, which fulfil the requirements to be included in our analysis. The green bars show the distribution for funds with positive alpha forecasts. The red bars show the distribution for funds with positive alpha forecasts and positive average predictive alpha. The blue bars show the distribution of funds included in the superior FCS assuming a narrow set ($\lambda = 0.90$) and using the time-varying latent skill and holdings performance model.
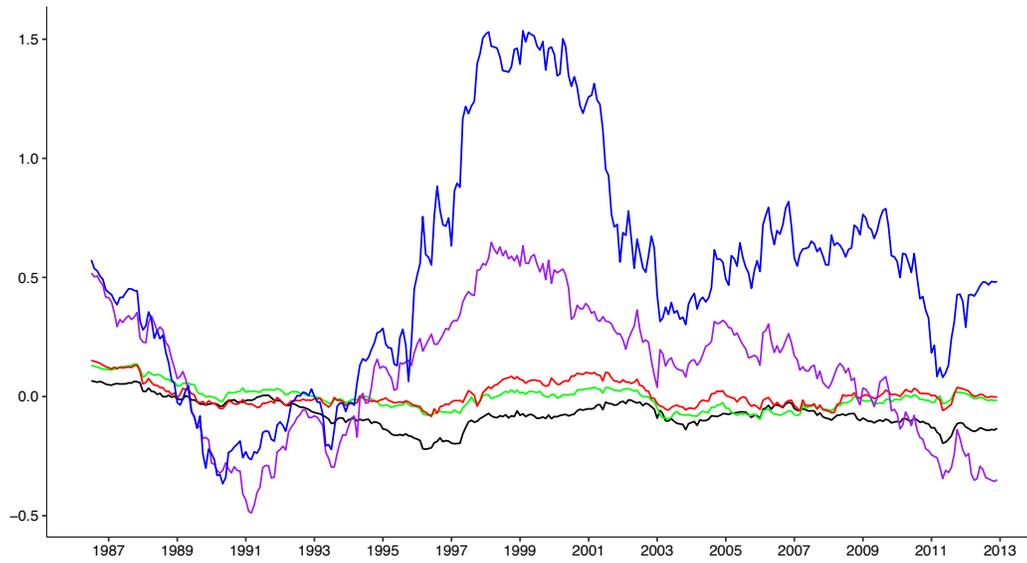
Figure 2: Time variation in alpha estimates for superior funds. This figure shows nonparametric estimates of local time-variation in four-factor alpha estimates (in percent per month) for a variety of portfolios consisting of top-ranked funds. In each portfolio the weight on fund i is given by $\frac{\hat{\alpha}_{i,t+1|t}/\hat{\sigma}_i^2}{\sum_{j=1}^{n}\hat{\alpha}_{j,t+1|t}/\hat{\sigma}_j^2}$. 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 forecasts of alpha while the red line tracks the performance of a portfolio consisting of the funds with positive alpha forecasts and positive average predictive alpha. The purple line tracks the performance of a portfolio consisting of the top 5% of predictive alpha-ranked funds with positive alpha forecasts and positive average predictive alphas. For comparison the black line tracks the variation in the alpha of the average fund. The sample period is 1986:07-2012:12.
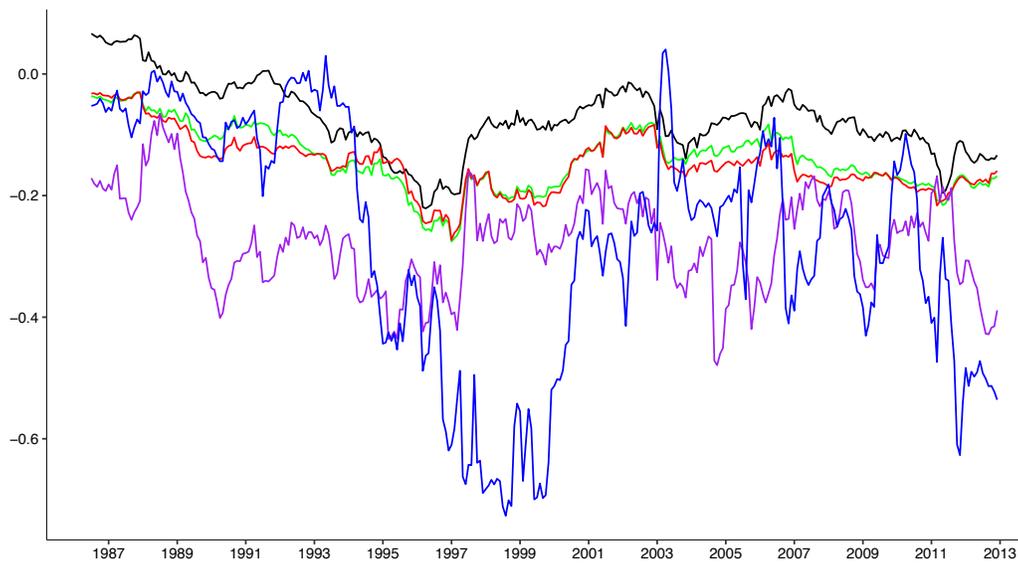
Figure 3: Time variation in alpha estimates for inferior funds. This figure shows nonparametric estimates of local time-variation in four-factor alpha estimates (in percent per month) for a variety of portfolios consisting of top-ranked funds. In each portfolio the weight on fund i is given by $\frac{\hat{\alpha}_{i,t+1|t}/\hat{\sigma}_i^2}{\sum_{j=1}^{n}\hat{\alpha}_{j,t+1|t}/\hat{\sigma}_j^2}$. 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 forecasts of alpha while the red line tracks the performance of a portfolio consisting of the funds with negative alpha forecasts and positive average predictive alpha. The purple line tracks the performance of a portfolio consisting of the bottom 5% of predictive alpha-ranked funds with negative alpha forecasts and positive average predictive alphas. For comparison the black line tracks the variation in the alpha of the average fund. The sample period is 1986:07-2012:12.
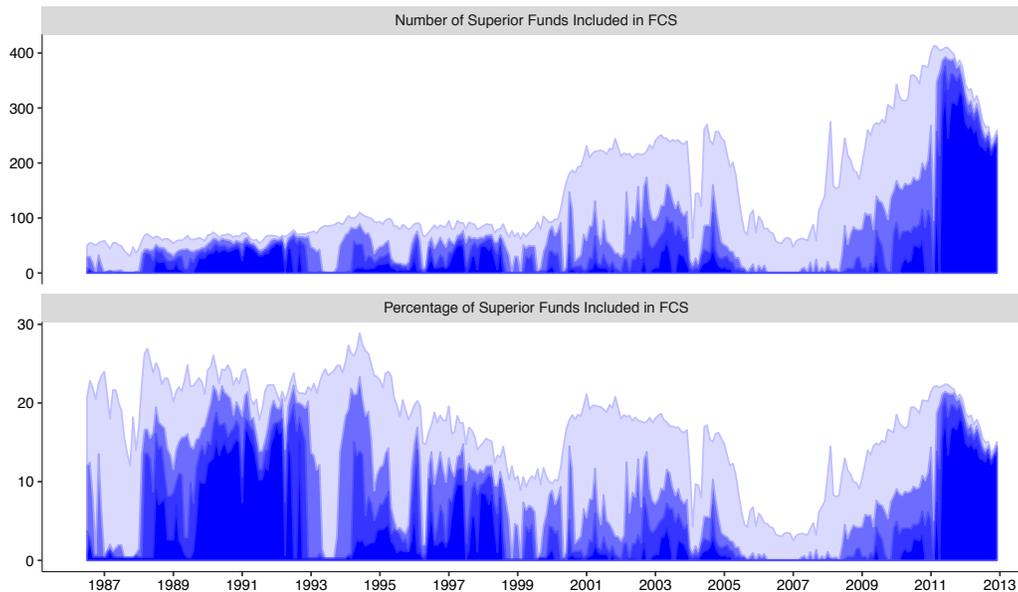
Figure 4: Evolution in the set of funds identified as being superior by the FCS approach. The top panel shows the evolution in the number of funds included in the tight, medium and wide fund confidence sets of superior funds. The bottom panel shows the corresponding evolution in the percentage of funds 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).
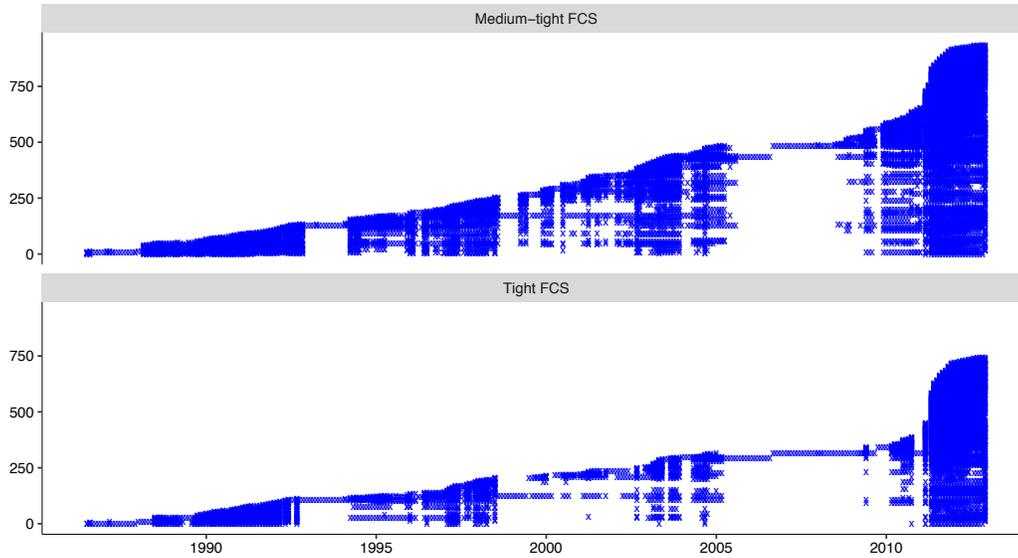
Figure 5: Identification of individual funds with superior skill by the FCS approach. The plot illustrates the evolution in the composition of individual funds with superior skills as identified by the FCS approach. Each fund that is included in the FCS at least once during the sample is assigned a unique 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.
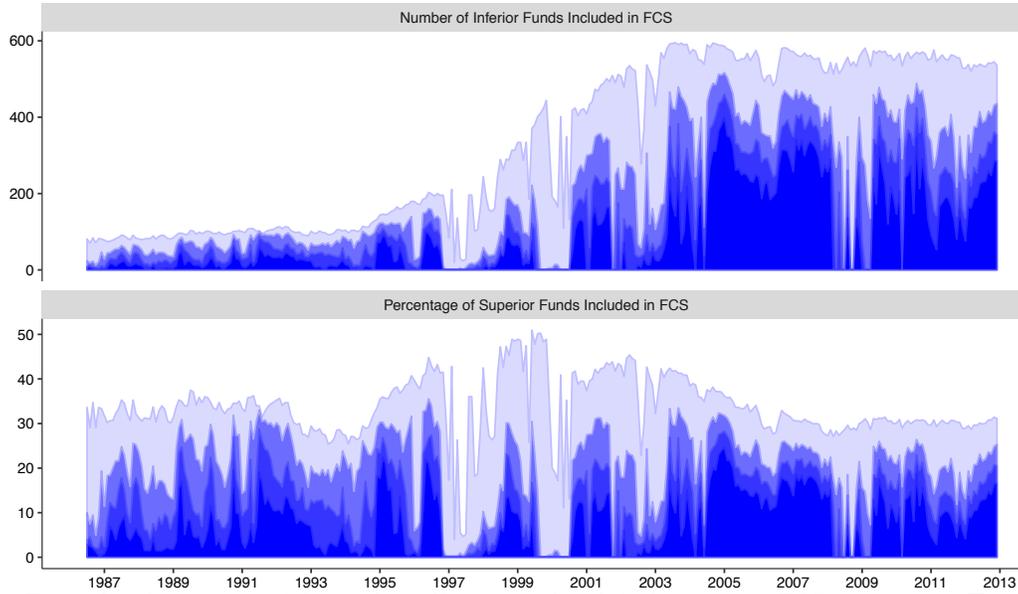
Figure 6: Evolution in the set of funds identified as being inferior by the FCS approach. The top panel shows the evolution in the number of funds included in the tight, medium and wide fund confidence sets of inferior funds. The bottom panel shows the corresponding evolution in the percentage of funds 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).
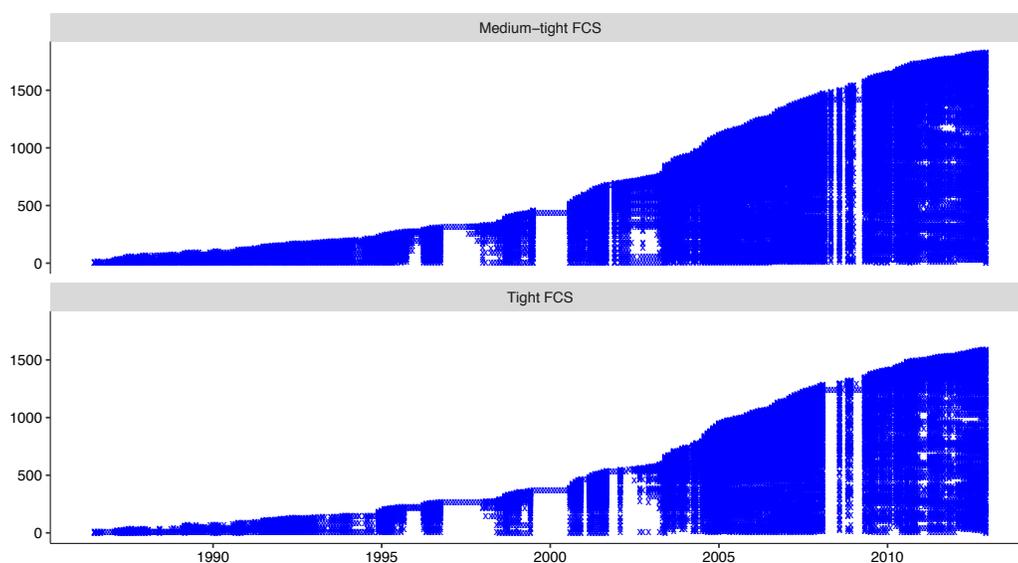
Figure 7: Identification of individual funds with inferior skill by the FCS approach. The plot illustrates the evolution in the composition of individual funds with inferior skills as identified by the FCS approach. Each fund that is included in the FCS at least once during the sample is assigned a unique 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.
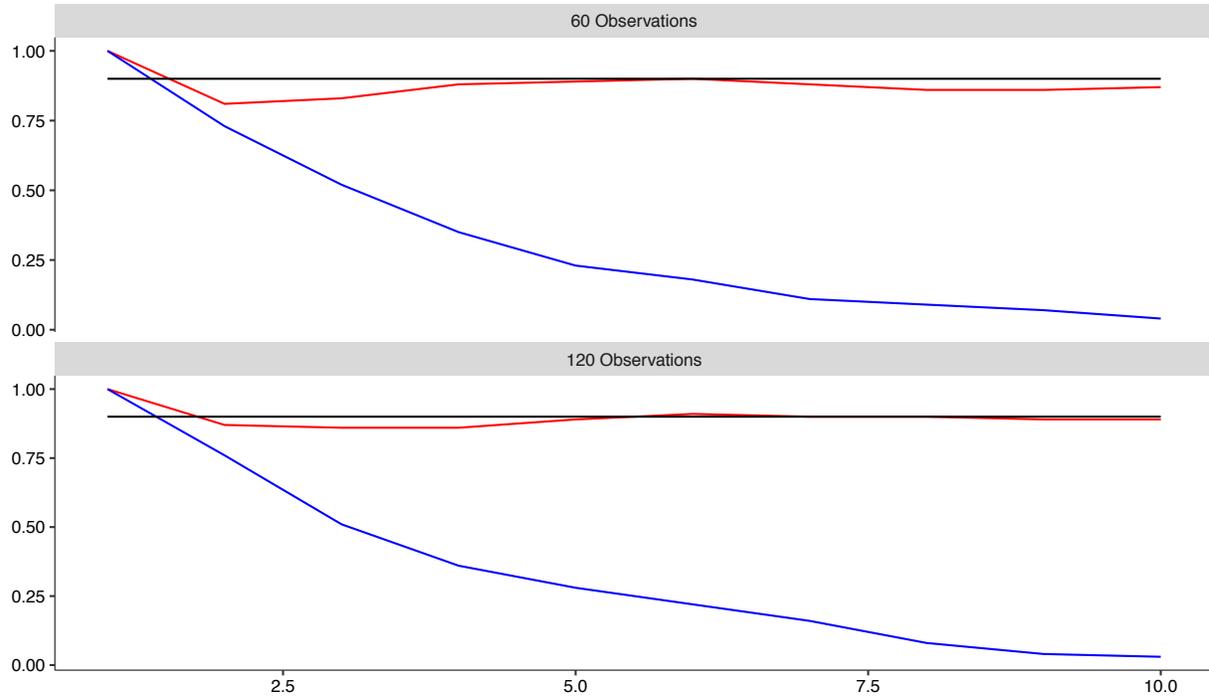
Figure B.1: The figure presents simulation results for the overall level of two selection methods. The red line represents the level of the FCS procedure, while the blue line represents the alternative specification based on the maximum t-statistic. The black line represents the nominal level of the tests. The level is presented as a function of the percentage of truly superior funds in the simulated sample of 100 funds. The levels are calculated as average rejection rates over 100 simulations. The top panel presents results for a sample size of 60 observations and the bottom panel presents results for a sample size of 120 observations.